

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: January 13, 2004, 05:14:36 ; Search time 4187 Seconds  
(without alignments)  
11383.109 Million cell updates/sec

file: US-09-922-225A-1

effect score: 1961

sequence: 1 agcgttcacgtgctagtcg.....aggtctctgacatgaaaaa 1961

coring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

searched: 22781392 seqs, 12152238056 residues

total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estc:\*

9: gb\_esti:\*

10: gb\_est2:\*

11: gb\_estc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_man:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1444.4	73.7	2036	11 AK054293	AK054293 Mus muscu
2	1442.8	73.6	2586	11 AK050335	AK050335 Mus muscu
3	1435	73.2	2591	11 AK086674	AK086674 Mus muscu
4	977.4	49.8	1201	9 AL516558	AL516558 AL516558

C	5	931.4	47.5	972	13	BX383922	BX383922
	6	879.6	44.9	1160	13	BX345346	BX345346
	7	869.2	44.5	1006	13	BX383923	BX383923
C	8	869.4	44.3	1201	9	AL516557	AL516557
	9	852.8	43.5	891	13	BX329238	BX329238
	10	850.4	43.4	1080	13	BX361838	BX361838
	11	816.4	41.6	933	13	BX369641	BX369641
C	12	813.6	41.5	942	9	AL576475	AL576475
	13	809.2	41.3	851	13	BU186063	BU186063
	14	802.8	40.9	1016	12	BM450315	BM450315
	15	791.6	40.4	856	13	BU132994	BU132994
	16	790.6	40.3	859	12	BM82816	BM82816
	17	790	40.3	1054	12	BM920754	BM920754
	18	787.8	40.2	837	12	BI092420	BI092420
	19	775.8	39.6	873	14	CD358784	CD358784
	20	774	39.5	1037	12	BM468168	BM468168
	21	772.4	39.4	853	14	CD300154	CD300154
C	22	749	38.2	771	13	BU739289	BU739289
	23	742	37.8	872	13	BU167220	BU167220
	24	730.4	37.2	824	12	BM456432	BM456432
	25	726	37.0	855	12	BI825443	BI825443
	26	722.4	36.8	946	10	EG674191	EG674191
	27	722.4	36.8	979	10	BG691522	BG691522
	28	722.2	36.8	1048	13	BQ955364	BQ955364
	29	717.6	36.6	1154	12	BM906750	BM906750
	30	713.4	36.4	960	10	BF795352	BF795352
	31	710	36.2	751	10	EG701766	EG701766
	32	700.6	35.7	1102	12	BM906325	BM906325
	33	687.2	35.0	924	12	BM451188	BM451188
	34	686.2	35.0	1120	12	BM805877	BM805877
	35	684.4	34.9	752	12	BI772096	BI772096
C	36	684	34.9	698	13	BU623771	BU623771
	37	681.6	34.8	775	10	EG717364	EG717364
	38	675.8	34.5	953	10	BF795135	BF795135
	39	673.2	34.3	939	10	BF795137	BF795137
	40	661.8	33.7	1182	12	BM545693	BM545693
	41	650.4	33.2	653	12	BM714659	BM714659
	42	648	33.0	713	10	EG717946	EG717946
	43	626.4	31.9	810	12	BG961872	BG961872
	44	626	31.9	682	14	CB155405	CB155405
	45	620.6	31.6	759	12	BM944500	BM944500

RESULT 1

AK054293

LOCUS

DEFINITION

Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN

full-length enriched library, clone:E33001F20 product:D1BD1

homolog [Homo sapiens], full insert sequence.

AK054293

VERSION

AK054293.1 GI:26344126

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374

PUBMED

11042159

ALIGNMENTS

AK054293 2036 bp mRNA linear HTC 05-DEC-2002  
Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN  
full-length enriched library, clone:E33001F20 product:D1BD1  
homolog [Homo sapiens], full insert sequence.

AK054293  
AK054293.1 GI:26344126  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.

1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
PUBMED  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
PUBMED  
11042159







844 TCCTGGTGGCCCTGGCTCTCTCTTTCTGGTGGCCGGTGGTCAATGACAGCTACTATTAC 903  
832 GGAAGTTGGTGATTCGACACCTCAACATCTGTTTGTATATAGTCTTTACTCCTCATGGA 891  
904 GGAAGTTGGTGGTGGCCCTCTCAACATCTGTTTGTATATAGTCTTTACTCCTCATGGA 963  
892 CTTGATCTTTATGGTACAGAACCTGGTATTTCTATTTAAATTAATGGAATTTCTGAATTC 951  
964 CTTGATCTTTATGGTACAGAACCTGGTATTTCTATTTAAATTAATGGAATTTCTGAATTC 1023  
952 AATGTAGCTTTGCTTTGGCTCTCTCTAGTCTCACTACCTGACTTCTTTATGGAATACCTG 1011  
1024 AATGTTGCCCTTTGGCTCTCTCTGGTCTTGGCTGCTGCTTCTTATGGAATACCTG 1083  
1012 CTGACAGATTTCAATGTTCAAGATTTAGGCAACCGTATGCTGCTTACCTTGGCTCCAATG 1071  
1084 CTGACAGATTTCAATGTTCAAGATTTAGGCAACCGTATGCTGCTTACCTTGGCTCCAATG 1143  
1072 TATATTTGGTTTATATTTCTTATCCAGCTCTCAAGAGAGAGAGATTTCTTTTCCT 1131  
1144 TATATTTGGTTTATATTTCTTATCCAGCTCTCAAGAGAGAGAGATTTCTTTTCCT 1203  
1132 GTGATATCAATATATGTTCTCTGCGGCTGCTGCTCTCTGCTGCTTCAAGAAATGTTAC 1191  
1204 GTCTATCCGCTCATATGCTCTCTGCTGCTGCTGCTCTCTGCTGCTTCAAGAAATGTTAC 1263  
1192 CACTTTGCTTTCAACGATATCGCTGGAGCACTACTGTCGACATCGAATTTGGCTGGCA 1251  
1264 CACTTTGCTTTCAACGATATCGCTGGAGCACTACTGTCGACATCGAATTTGGCTGGCA 1323  
1252 TTAGGAATCTGCTTTCCGTTTGGCTCTTGTCTATTTCTGCTGCTGCTGCTGCTGCTGCT 1311  
1324 CTGGGACAGTTTTCCTGTTTGGCTCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1383  
1312 GGATATCAAGGCGCCCTGTTGTTATCCAGATTTTACGATTTGCTACAGACCCCAACC 1371  
1384 GGATATCAAGGCGCCCTGTTGTTATCCAGATTTTACGATTTGCTACAGACCCCAACC 1443  
1372 ATCCACACTGTCTCCAGAGGAGCAGCTGTGAATGCTGTGTGGGAAAGAGTGGTATCGA 1431  
1444 ATCCACACTGTCTCCAGAGGAGCAGCTGTGAATGCTGTGTGGGAAAGAGTGGTATCGA 1503  
1432 TTTCGAGAGCTTCTCTCTCTGACAAATGTCAGCTTCACTTCACTCCATCAGAGTTC 1491  
1504 TTTCGAGAGCTTCTCTCTCTGACAAATGTCAGCTTCACTTCACTCCATCAGAGTTC 1563  
1492 AGAGTCACTTACCAAAACCTTTTGCAGAGGACCTCTGCGCCACCGGATTTGTTCTACT 1551  
1564 AGAGTCACTTACCAAAACCTTTTGCAGAGGACCTCTGCGCCACCGGATTTGTTCTACT 1623  
1552 GACATGATGACAGATCTAGAGAGCCATCCAGATATATGATATGATTAATGATTAATGCCAT 1611  
1624 CACATGATGACAGATCTAGAGAGCCATCCAGATATATGATTAATGATTAATGCCAT 1683  
1612 TATTTAGTGGATTTGGACACCATCAGAGAAACCCCGGAGGACCAAAATATTCATCCAT 1671  
1684 TACTTAGTGGATTTGGATACCATCAGAGAAACCCCGGAGGACCAAAATATTCATCCAT 1743  
1672 AAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1731  
1744 AAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1803  
1732 CTGCTGCGGATTTCTATGTTCCCTTCTGTCAGATCAGTATCAGTATCAGTATCAGTATCAG 1791  
1804 TTGCTTGGGATTTCTATGTTCCCTTCTGTCAGATCAGTATCAGTATCAGTATCAGTATCAG 1863  
1792 ACCATCTCAAAACCCCGGAGGACCAAAATGATGATGATGATGATGATGATGATGATGAT 1851  
1864 ACCATCTCAAAACCCCGGAGGACCAAAATGATGATGATGATGATGATGATGATGATGAT 1920  
1852 ACCTGTCGCGGACCAAGGACCAATCTGTTGTTTCACTTATGATTTCCAGTACCTGACCTCCCT 1911

Db 1921 CCCGTGACTCCAGAGGACGCC--TGTGCTAACTAGGGCTTCCACGGAGCCGACTTCCT 1978  
Qy 1912 GCAAGTCATCGCTTAACATTTGTATTAAGAGTCTTCTGACATGA 1957  
Db 1979 CCAAGGCAT--CCTGTTAGAGCTGTATTAAGAGTCTGCTCCACCAAGA 2022

RESULT 3  
AK086674 2591 bp mRNA linear HTC 05-DEC-2002  
LOCUS Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:D330045C02 product:D19D1 homolog [Homo sapiens], full insert sequence.  
DEFINITION  
ACCESSION AK086674.1 GI:26352160  
VERSION HTC; CAP trapper.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watanabe, M., Yoneda, S., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, S., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bolunga, N., Toyooka, K., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamly, M., Lee, N. H., Lyons, P., Marchionni, B., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Rongwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. P., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilmberg, L., Wyshew-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohtsuki, S. and Hayashizaki, Y.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409 (6821), 685-690 (2001)  
MEDLINE 21085660  
PUBMED 11217851  
REFERENCE 5  
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research



```
1132 GTGTATCCACTATATGTCTCTGTGGCGTGTGGCTCTCTGTGCACTTCAGAAATGTTAC 1191
1223 GTCTATCCGCTCATATGCTCTGTGGTGTGTGTCTCTCTGTGCACTTCAGAAATGTTAC 1282
1192 CATTGTGTGTTTCAAGATATCGCTGGAGCACTATCTGTGACATCGAATGCTGGCA 1251
1283 CATTGTGTGTTTCAAGATATCGCTGGAGCACTATCTGTGACATCGAATGCTGGCG 1342
1252 TTAGGAAGTCTCTCTCTGTGTTGGCTCTCTGTGATTTCTCTGCTGTGGCACTGTTCAGA 1311
1343 CTGGGACAGCTTTCTCTGTGTTGGCTCTCTGTGATTTCTCTGCTGTGGCTGTTCAGA 1402
1312 GATATACACGGGCGCTCTGTGTTGGCTCTCTGTGATTTCTCTGCTGTGGCACTGTTCAGA 1371
1403 GGATACACGGGCGCTCTGTGATTTCTCTGATTTATATCCAGAAATTTTACAGAAATGGCCACACCCAAACC 1462
1372 ATCCACACTCTCCAGAGGAGCAGACCTGTGAATGTCTGTGGGAAAAGAGTGGTATCGA 1431
1463 ATCCACACTGTCCAGAGGAGCAGGCGGTGAAGTCTGTGTGGGAAAAGAGTGGTATCGA 1522
1432 TTTCAGAGAGCTTCTCTCTGTGATTTGGCAATTTGGCACTTCAGTTCATTCATCCAGAGTTC 1491
1523 TTTCAGAGAGTCTCTCTCTGTGATTTGGCAATTTGGCACTTCAGTTCATTCATCCAGAGTTC 1582
1492 AGAGTCTAGTATACCAAAACCTTTTGAGAGGAGCTCTGGCCACCGGATTTGTTCTTACT 1551
1583 AGAGTCTAGTATACCAAAACCTTTTGAGAGGAGCTCTGGCCACCGGATTTGTTCTTACT 1642
1552 GACATGAATGACAGAAATCTAGAGAGGCAATCCAGATATTTATGATATCGATTAATGCCAT 1611
1643 CACATGAACAGCAGAAACCGAGAGGAGCTTCCAGATATTTATGATATCGATTAATGCCAT 1702
1612 TATTAGTGTGATTTGGACACCATGAGAGAAACCC - CGGAGGAGCAAAATATTCATCAA 1670
1703 TACTTGTGATTTGGATACATGAGAGAAACCATCTGGAGGCAAACTACTCATCCA 1762
1671 TAAAGAGATGATGATGATGCTTGGCTATGACATTCATTCCTTGTGATCTTGTAGATCTTCAA 1730
1763 CAGAGAGAGTGGGTGAGCTGCTGCTGACAGACATTTCTTGGATGCTGTAGGCTTCCAA 1822
1731 GCTCTGGGCGATTTCTATGTCCTTCTGTCAGATCAGTATACAGTGTAGTAACTA 1790
1823 GTTCTTGGGCTATTTATGTTTCTCTCTGTCAGATCAGTATACAGTGTATGTGACTA 1882
1791 CACATCTCAAACCCCGAAGCAAAAGCAAAATCAGAGAAAGAGTGGAGTTAGCAACA 1850
1883 CACATCTGAAGCTCGGAAAGCAAAAGCGAGCAGAGAAAGAGTGGAGTTAG---CA 1939
1851 CACCTGTGGCGCCCAAGAGCAACCATCTGTTAACTATTGATTTCCAGTCACTGACTCC 1910
1940 CCCCGGTGACTCCAGAGGAGAGCC--TGTCCTAACTAGGCTTCCACGGAGCGGACTTCC 1997
1911 TGCAAGTCACTGCTGTAAACATTTGTAATAAAGTCTTCTGACATCA 1957
1998 TCCAAGGAT--CTGTGTAGAGTGTATAAAGGCTGTCCACCAAGA 2042
```

```
RESULT 4
AL516558 1201 bp mRNA linear EST 09-MAY-2003
JOCUS
DEFINITION AL516558 Homo sapiens NEUROBLASTOMA Homo sapiens cdna clone
CSODA006YN06 5-PRIME, mRNA sequence.
ACCESSION AL516558
VERSION AL516558.2 GI:30491310
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li W.B., Gruber C., Jeesee J. and Polayes D.
Full-length cdna libraries and normalization
JOURNAL unpublished
```

## COMMENT

On Feb 13, 2001 this sequence version replaced gi:12780051.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
Bp 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5955.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODA006D03QPI&cluster=5955.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CSODA006D03QPI.

FEATURES  
source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODA006YN06"  
/tissue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cdna was primed  
with a NotI-oligo(dT) primer. Five prime end enriched into  
double-strand cdna was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 247 a 257 c 274 g 363 t 60 others  
ORIGIN

Query Match 49.8%; Score 977.4; DB 9; Length 1201;  
Best Local Similarity 94.1%; Pred. No. 2,1e-225;  
Matches 1027; Conservative 35; Mismatches 22; Indels 7; Gaps 5;

```
QY 10 ATGCTAGTTCGAGGGGCTCGGAGCGCTGGAAGGCGAGCGGGCCAGCAGTGGGGATACG 69
DB 50 ATGCTAGTTCGAGGGGCTCGGAGCGCTGGAAGGCGAGCGGGCCAGCAGTGGGGATACG 109
QY 70 GCGCGGCTGCGGACAAAGCTGCGGGAGCTGCTGGGCGAGCGGAGCGGGCGGCGGAG 129
DB 110 GCGCGGCTGCGGACAAAGCTGCGGGAGCTGCTGGGCGAGCGGAGCGGGCGGCGGAG 169
QY 130 CACCGAGCGAGTATCTGGGAAACAAAGCAGGACAAAGTCTGGGCACTCTGAAGGATCTACT 189
DB 170 CACCGAGCGAGTATCTGGGAAACAAAGCAGGACAAAGTCTGGGCACTCTGAAGGATCTACT 229
QY 190 GCTTCAAGTGTCTGTTTTCAGCAAGGTTATGTGCTGCTCTCTGAGCAACATCTCTGAC 249
DB 230 GCTTCAAGTGTCTGTTTTCAGCAAGGTTATGTGCTGCTCTCTGAGCAACATCTCTGAC 289
QY 250 TGTGATGAACATTTCAACTACTGGGAGCCACACACTACCTCATCTATGGGAAAGGGTTT 309
DB 290 TGTGATGAACATTTCAACTACTGGGAGCCACACACTACCTCATCTATGGGAAAGGGTTT 349
QY 310 CAGACTTGGGAATATTTCCCGCAGCATATGCCATTTCGCTCTATGTCTTACCTGTCTTCAAT 369
DB 350 CAGACTTGGGAATATTTCCCGCAGCATATGCCATTTCGCTCTATGTCTTACCTGTCTTCAAT 409
QY 370 GCGTGGCGAGCTGCATTTTCATCGAAGATTTTACAACTTAATAGATCTTGTGTTTAC 429
DB 410 GCGTGGCGAGCTGCATTTTCATCGAAGATTTTACAACTTAATAGATCTTGTGTTTAC 469
QY 430 TTTTGTGCGATGCTTCTGCGCTTTTGTGAGCTGTATTGTGAACTTTTACTTTTACAAGCT 489
DB 470 TTTTGTGCGATGCTTCTGCGCTTTTGTGAGCTGTATTGTGAACTTTTACTTTTACAAGCT 529
QY 490 GTGTGCGAAGAGTTTGGTGTGACGTGAGTCAAAATGATGCTAGCTTCTTGGTCTCAGC 549
DB 530 GTGTGCGAAGAGTTTGGTGTGACGTGAGTCAAAATGATGCTAGCTTCTTGGTCTCAGC 589
QY 550 ACTGGGATGTTTTGCTCATCATCAGCAATTCCTTCTCTAGTAGTCTTCTGTATGTACTACG 609
DB 590 ACTGGGATGTTTTGCTCATCATCAGCAATTCCTTCTCTAGTAGTCTTCTGTATGTACTACG 649
```



```

610 TTGATAGCCATGACTGGATGATATATGACCAAGACTTCCATTGCTGTGCTGGAGTAGCA 669
650 TTGATAGCCATGACTGGATGATATATGACCAAGACTTCCATTGCTGTGCTGGAGTAGCA 709
670 GTGGGGGTACTTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729
710 GTGGGGGTACTTATGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 769
730 TTGCTGGTCAATGAACACAGAGTGAAGAGTTCCTTCAATGGTGGCTGATGGCCCTCAT 789
770 TTGCTGGTCAATGAACACAGAGTGAAGAGTTCCTTCAATGGTGGCTGATGGCCCTCAT 829
790 CTATTTCTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 849
830 CTATTTCTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 889
850 CCACCTCAACATTTTGTATATAATGCTCTTACTCCTCATGACCTGATCTTTATGATGA 909
890 CCACCTCAACATTTTGTATATAATGCTCTTACTCCTCATGACCTGATCTTTATGATGA 949
910 GAACCTGGTATTTCTATTAATTAATGATTTCTGATTTCAATGATGAGCTTTGCTTTG 969
950 GAACCTGGTATTTCTATTAATTAATGATTTCTGATTTCAATGATGAGCTTTGCTTTG 1009
970 GCTCTCTAGTCTCTACCTGACTCTCTCTTATGAATACCTGCTGACAGATTTGATTT 1029
1010 GCTCTCTAGTCTCTACCTGACTCTCTCTTATGAATACCTGCTGACAGATTTGATTT 1066
1030 GAGATTTAGGACCCGATTTAGGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1089
1067 GAGATTTAGGACCCGATTTAGGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1122
1090 TTCTCTCATCCA 1100
1123 TVAHCMCMCCM 1133

```

```

RESULT 5
X383922/c
CCUS
EFINITION
CESSION
ERSON
EWORDS
OURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
X383922 Homo sapiens HELA CELLS COT 25-NORMALIZED EST 08-MAY-2003
cDNA clone CS0DK002YM06 3-PRIME, mRNA sequence.
X383922
X383922.1 GI:30459192
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 972)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK002BG03NP1&cluster=5955.r. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DK002BG03NP1.
Location/Qualifiers
1..972
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YM06"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
FEATURES
source

```

```

/clone.lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 269 a 196 c 242 g 247 t 18 others
ORIGIN
Query Match 47.5%; Score 931.4; DB 13; Length 972;
Best Local Similarity 98.3%; Pred. No. 2.7e-214;
Matches 943; Conservative 6; Mismatches 9; Indels 1; Gaps 1;
QY 974 TCCTAGTCTTACCACTGACCTTCTCTTATGGAATACCTGCTGACAGAGATTTCATGTTTCA 1033
DB 972 TCCTAGTCTTACCACTGACCTTCTCTTATGGAATACCTGCTGACAGAGATTTCATGTTTCA 913
QY 1034 ATTTAGGCCACCGTATGCTTACCTTGGCTCCATGATATTTGGTTTATAAATTTCT 1093
DB 912 ATTTAGGCCACCGTATGCTTACCTTGGCTCCATGATATTTGGTTTATAAATTTCT 853
QY 1094 TCATCCAGCCTCACAAGAGAGAGAGATTTCTTTTCCCTGCTATCCACTTATATGTTCTCT 1153
DB 852 TCATCCAGCCTCACAAGAGAGAGAGATTTCTTTTCCCTGCTATCCACTTATATGTTCTCT 793
QY 1154 GTGGCGTGTGCTCTCTCTGCACTTCAGAAATGTTTACCACTTGTGTTTCAACGATATC 1213
DB 792 GTGGCGTGTGCTCTCTCTGCACTTCAGAAATGTTTACCACTTGTGTTTCAACGATATC 733
QY 1214 GCCTGGAGCACTATACCTGACATCGAATTTGGCTGCAATAGGAATGCTTCTTCTGTTTG 1273
DB 732 GCCTGGAGCACTATACCTGACATCGAATTTGGCTGCAATAGGAATGCTTCTTCTGTTTG 673
QY 1274 GGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTACAGAGATATCAGGGC-CCCTTGAT 1332
DB 672 GGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTACAGAGATATCAGGGCACCCTTGAT 613
QY 1333 TTGTATCCAGAAATTTTACCGAATTTGTACAGACCCCAACACTCCACACTGTCCAGAGAGC 1392
DB 612 TTGTATCCAGAAATTTTACCGAATTTGTCTACAGACCCCAACACTCCACACTGTCCAGAGAGC 553
QY 1393 AGACTGTGAATGCTCTGTGGGAAAAGAGTGTATCGAATTTCCAGAGAGCTTCTTCTTCT 1452
DB 552 AGACTGTGAATGCTCTGTGGGAAAAGAGTGTATCGAATTTCCAGAGAGCTTCTTCTTCT 493
QY 1453 CCTGACAAATGGCAGCTTTCAGTTCATTCATCAGAGTTCAGAGGTCAGTTTACCAAACT 1512
DB 492 CCTGACAAATGGCAGCTTTCAGTTCATTCATCAGAGTTCAGAGGTCAGTTTACCAAACT 433
QY 1513 TTTGAGAGAGCACTCTGGCCACCCGAGTTGTTTCTTACTGACATGAATGACCAAGATCTA 1572
DB 432 TTTGAGAGAGCACTCTGGCCACCCGAGTTGTTTCTTACTGACATGAATGACCAAGATCTA 373
QY 1573 GAAGAGCACTCCAGATATATGATATCAGTAATGCAATTTAGTGGATTTGGACACC 1632
DB 372 GAAGAGCACTCCAGATATATGATATCAGTAATGCAATTTAGTGGATTTGGACACC 313
QY 1633 ATGAGAGAAAACCCCGGAGCCAAAATATTTCATCCAAATAAGAAGATGATGATCAGCTTG 1692
DB 312 ATGAGAGAAAACCCCGGAGCCAAAATATTTCATCCAAATAAGAAGATGATGATCAGCTTG 253
QY 1693 GCCTATAGCACTTCTTGTGCTTCTAGATCTTCAAGCTGCTGGGCACTTCTATGTC 1752
DB 252 GCCTATAGCACTTCTTGTGCTTCTAGATCTTCAAGCTGCTGGGCACTTCTATGTC 193
QY 1753 CCCTTCTCTGTCAGATCAGTATACAGTGTAGTAACCTTACCACTCTCTCAAAACCCGAAA 1812
DB 192 CCCTTCTCTGTCAGATCAGTATACAGTGTAGTAACCTTACCACTCTCTCAAAACCCGAAA 133
QY 1813 GCAAGCAAAATCAGAGAAAAGTGGAGGTAGCAACACACCTGTGCGCCCAAGAGCAA 1872
DB 132 GCAAGCAAAATCAGAGAAAAGTGGAGGTAGCAACACACCTGTGCGCCCAAGAGCAA 73
QY 1873 CCATCTTGTAACTATTGATTTCCAGTGCCTGACTCTCCCTGCAAGTCAATCGCTGTAA 1931

```



```

db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| : : :
72 CCATCTTTTACATTGATTCAGTAGCACTGCGCCCTGCANGTCTGCCCGCWNV 14

BX345346                               1160 bp      mRNA       linear    EST 02-MAY-2003
BX345346 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
CDNA clone CS0DC013YC23 5-PRIME, mRNA sequence.
BX345346
BX345346.1 GI:30347354
EST.
Homo sapiens (human)
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1160)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: secre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5954.r For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AAW192G03RM1&cluster=5954.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0AAW192G03RM1.
Location/Qualifiers
1..1160
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC013YC23"
/tissue type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note=Tst strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized.
BASE COUNT          251 a   269 c   283 g   353 t     4 others
ORIGIN
Query Match              44.9%; Score 879.6; DB 13; Length 1160;
Best Local Similarity   97.3%; Pred. No. 9.7e-202;
Matches 905; Conservative 0; Mismatches 24; Indels 1; Gaps 1;
2y 10 ATGGCTAGTCGAGGGCTCGGCAGCGCTGAGGGCAGCGGGGCCAGACTGGGATACG 69
Db 60 ATGGCTAGTCGAGGGCTCGGCAGCGCTGAAGGGCAGCGGGGCCAGACTGGGATACG 119
2y 70 GCCCGGGCTCGGACAAGCTCGGGGAGCTGTGGGACGCGAGAGCGGGCGCGCGAG 129
Db 120 GCCCGGGCTCGGACAAGCTCGGGGAGCTGTGGGACGCGAGCGGGCGCGCGAG 179
2y 130 CACCGGACCGAGTTATCTGGGAACAAGACGAGCAAGCTCGGGCACCTGAGAGTCTACT 189
Db 180 CACCGGACCGAGTTATCTGGGAACAAGACGAGCAAGCTCGGGCACCTGAGAGTCTACT 239
2y 190 GCTTTCAAGTGTCGTTTTCAAGAGTTATGTGCTGCTCTCTCTGAGCAACATCTCTGAC 249
Db 240 GCTTTCAAGTGTCGTTTTCAAGAGTTATGTGCTGCTCTCTCTGAGCAACATCTCTGAC 299
2y 250 TGTGATGAACATTCAACTACTGGGAGCAACACACTTCACTCTATGGGAGGGTTT 309
Db 300 TGTGATGAACATTCAACTACTGGGAGCAACACACTTCACTCTATGGGAGGGTTT 359
2y 310 CAGACTTGGGAATATTCGCCAGCATATGCCATTGCTCTCTATGTTACCTGTTGTTCTAT 369

```

Query Match	44.5%	Score	872.2	DB	13	Length	1006	
Best Local Similarity	99.4%	2	Mismatches	1	Indels	2	Gaps	2
Matches	893	Conservative						
y	1	AGGCTTTGCCATGGCTAGTCGAGGGGCTCGCGAGCCCTGAAGGGCAGCGGGGCCAGCAGT	60					
b	111	AGGCTTTGCCATGGCTAGTCGAGGGGCTCGCGAGCCCTGAAGGGCAGCGGGGCCAGCAGT	170					
y	61	GGGGATACGCCCCCGGCTCGGGAACAAGCTCGCGAGCTCTCGGGCAGCGAGAGCGCGG	120					
b	171	GGGGATACGCCCCCGGCTCGGGAACAAGCTCTCGGGAGCTCTCGGAGCGAGAGCGGCG	230					
y	121	GGGCGGAGCACCGGACCGAGTTATCTGGGAAACAAAGCAGGCAAGCTCTGGGCACTCTGAA	180					
b	231	GGGCGGGA-SACCGGACCGAGTTATCTGGGAAACAAAGCAGGCAAGCTCTGGGCACTCTGAA	289					
y	181	GGATCTACTGCTTTCAAGTCTCTGCTTTTCAGCAAGGTTATGTGCTGCTCTCTCGAGCAAC	240					
b	290	GGATCTACTGCTTTCAAGTCTCTGCTTTTCAGCAAGGTTATGTGCTGCTCTCTCGAGCAAC	349					
y	241	ATCTCTGACTGTGATGAAACATTCAACTACTGGAAGCCAAACAATACCTCATCTATGGG	300					
b	350	ATCTCTGACTGTGATGAAACATTCAACTACTGGAAGCCAAACAATACCTCATCTATGGG	409					
y	301	GAAGGGTTTCAGACTCTGGGAATATTCCCAAGCATATGCCATTTCGCTCTGCTTACCTG	360					
b	410	GAAGGGTTTCAGACTCTGGGAATATTCCCAAGCATATGCCATTTCGCTCTGCTTACCTG	469					
y	361	TTGCTTCATGCTGGCCAGCTGCATTTTCATGCAAGAAATCTACAAACTAATAGATTCTT	420					
b	470	TTGCTTCATGCTGGCCAGCTGCATTTTCATGCAAGAAATCTACAAACTAATAGATTCTT	529					
y	421	GTGTTTTTACTTTTTTGGCATGTCTTCTGGCTTTTGTGAGCTGTATTTGTGAACCTTTACTTT	480					
b	530	GTGTTTTTACTTTTTTGGCATGTCTTCTGGCTTTTGTGAGCTGTATTTGTGAACCTTTACTTT	589					
y	481	TACAAGGCTGTGTCCAGAAGTTTGGGTTGCACGTGATGCGAATGATGCTAGCCTCTTG	540					
b	590	TACAAGGCTGTGTCCAGAAGTTTGGGTTGCACGTGATGCGAATGATGCTAGCCTCTTG	649					
y	541	GTTCTCAGCACTGGCCATGTTTTTGTCTCATCATCAGCATTCCTCTCCTAGTACTTCTGATG	600					
b	650	GTTCTCAGCACTGGCCATGTTTTTGTCTCATCATCAGCATTCCTCTCCTAGTACTTCTGATG	709					
y	601	TACACTACGTTGATAGCCATGACTGGATGGTATATGACAAGACTTCCATGCTGTCGCTG	660					
b	710	TACACTACGTTGATAGCCATGACTGGATGGTATATGACAAGACTTCCATGCTGTCGCTG	769					
y	661	GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTCAGTCAGCTCTTGGTTTACCATT	720					
b	770	GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTCAGTCAGCTCTTGGTTTACCATT	829					
y	721	GCCCTTGATTTGCTGGTCAACAAACACAGGTGGGAAGAGTTTCTTTCATTCGTCGCTGATG	780					
b	830	GCCCTTGATTTGCTGGTCAACAAACACAGGTGGGAAGAGTTTCTTTCATTCGTCGCTGATG	889					
y	781	GCCCTCACTATTCTTGGTGCCTGTGGTGTCAATGACAGCTACTATATATGGGAAGTTG	840					
b	890	GCCCTCACTATTCTTGGTGCCTGTGGTGTCAATGACAGCTACTATATATGGGAAGTTG	949					

```
Db      890 TACCTTGGCTCCAAATGATATATTTGGTTTATATTTCTTCATCCAGCCTCACAAAGAGGA 831
Qy      1116 GAGATTTCTTTCCCTGCTGTATCCACTATATATGTCCTGTGGGCGCTGTGCTCTCTCTGC 1175
Db      830 GAGATTTCTTTCCCTGCTGTATCCACTATATATGTCCTGTGGGCGCTGTGCTCTCTCTGC 771
Qy      1176 ACTT-----CAGAAATGTTACCACTTTTGTG-TTTCACAGGATATC 1213
Db      770 ACTTCAGWACAGTTTCTGTACTTCCAGAAATGTTACCACTTTTGTGTTTTCACAGGATATC 711
Qy      1214 GCCTGGAGCATTATCTGTGATCATGAATTTGGCTGGGATTTAGGAACATGCTTCTCTGTTTG 1273
Db      710 GCCTGGAGCATTATCTGTGATCATGAATTTGGCTGGGATTTAGGAACATGCTTCTCTGTTTG 651
Qy      1274 GCCTCTTGTCATTTCTCGCTCTGTGGCACTGTTTCAGAGGATATCACGGGCCCTTGATT 1333
Db      650 GCCTCTTGTCATTTCTCGCTCTGTGGCACTGTTTCAGAGGATATCACGGGCCCTTGATT 591
Qy      1334 TGTATCCAGAAATTTACCGAATTTGCTACAGACCCAAACATCCACACTGTGCCAGAGGCA 1393
Db      590 TGTATCCAGAAATTTACCGAATTTGCTACAGACCCAAACATCCACACTGTGCCAGAGGCA 531
Qy      1394 GACCTGTGAATCTGTGTGGGAAAGAGTGTATCGAATTTCCAGCAGCTTCTCTCTTC 1453
Db      530 GACCTGTGAATCTGTGTGGGAAAGAGTGTATCGAATTTCCAGCAGCTTCTCTCTTC 471
Qy      1454 CTGACAAATGGCAGTTTCAGTTTCATTCATCAGAGTTTCAGAGGTCAGTTACCAAAACCTT 1513
Db      470 CTGACAAATGGCAGTTTCAGTTTCATTCATCAGAGTTTCAGAGGTCAGTTACCAAAACCTT 411
Qy      1514 TTGCGAGAGGACCTCTGGCCACCCGGATGTTCTCTACTGATCAATGAATGACCAAGTCTAG 1573
Db      410 TTGCGAGAGGACCTCTGGCCACCCGGATGTTCTCTACTGATCAATGAATGACCAAGTCTAG 351
Qy      1574 AAGAGCCATCCAGATATATTTGATATCAGTAAATGCCATTTATTTAGTGGATTTGGACACCA 1633
Db      350 AAGAGCCATCCAGATATATTTGATATCAGTAAATGCCATTTATTTAGTGGATTTGGACACCA 291
Qy      1634 TGAGAGAACACCCCGGAGCCAAATATTCATCCATAAGAGATGATCAGCTTGG 1693
Db      290 TGAGAGAACACCCCGGAGCCAAATATTCATCCATAAGAGATGATCAGCTTGG 231
Qy      1694 CCTATAGACCATTCCTTGATGCTTCTAGATCTTCAAAGCTGCTCGGGCATTCTATGTCC 1753
Db      230 CCTATAGACCATTCCTTGATGCTTCTAGATCTTCAAAGCTGCTCGGGCATTCTATGTCC 171
Qy      1754 CTTTCTGTGATCAGTATACAGTGTACGTAACTACACCATCTCAAACCCCGGAAAG 1813
Db      170 CTTTCTGTGATCAGTATACAGTGTACGTAACTACACCATCTCAAACCCCGGAAAG 111
Qy      1814 CAAAGCAATCAGGAAGAAAGTGGAGGTTAGCAACACACCTGTGGCCCCAAAGGACAAAC 1873
Db      110 CAAAGCAATCAGGAAGAAAGTGGAGGTTAGCAACACACCTGTGGCCCCAAAGGACAAAC 51
Qy      1874 CATCTTGTAACTA-TTGATTTCCAGTGTGACCTGATCCCTGCA 1914
Db      50 CATCTTGTAACTANNNTANTNCCAGTGACCTGACTSCBNVAA 9
```

```
RESULT 9
LOCUS   BX329238/c
DEFINITION BX329238 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
          CDNA clone CS0DK002YM06 3-PRIME, mRNA sequence.
ACCESSION BX329238
VERSION   BX329238.1 GI:30340822
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
```

```
TITLE
JOURNAL
COMMENT
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAI036ZC03_CS03379_1&cluster=5955.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAI036ZC03_CS03379_1.
FEATURES
Location/Qualifiers
1..891
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YM06"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 314 a 193 c 190 g 193 t 1 others
ORIGIN
Query Match 43.58; Score 852.8; DB 13; Length 891;
Best Local Similarity 99.34; Pred. No. 2.7e-195;
Matches 887; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
Qy 233 TGAGCAACATCTCTGACT-GTGATGAACATTCACACTACTGGGAGCCAAACACACTACCTC 291
Db 891 TGAGCAACATCTCTGACTGGTGTGATGAACATTCATCTCTGGGAGCCAAACACACTACNTC 832
Qy 292 ATCTATGGGAGAGGTTTCAGACTTGGGATATTCGCCAGCATATGCCATTCGCTCCTAT 351
Db 831 ATCTATGGGAGAGGTTTCAGACTTGGGATATTCGCCAGCATATGCCATTCGCTCCTAT 772
Qy 352 CTTTACCTGTTTCTCATGCTGCGCCAGCTGCATTTTCATGCAAGAAATTCACAACTAT 411
Db 771 CTTTACCTG-TGCTTCATGCTGCGCCAGCTGCATTTTCATGCAAGAAATTCACAACTAT 713
Qy 412 AAGATTCCTGTTGTTTACTTTTGGCATGTCCTCTGGCTTTTGTGAGCTGTATTGTGAA 471
Db 712 AAGATTCCTGTTGTTTACTTTTGGCATGTCCTCTGGCTTTTGTGAGCTGTATTGTGAA 653
Qy 472 CTTTACTTTTACAAGGCTGTGTGCAAGAAAGTTTGGGTTTCAGCTGAGTCGAATGATCTA 531
Db 652 CTTTACTTTTACAAGGCTGTGTGCAAGAAAGTTTGGGTTTCAGCTGAGTCGAATGATCTA 593
Qy 592 GCCTTCTGTTTCTCAGCACTGGCATGTTTGTCTCATCATCATGATTCCTCTCTAGTAGC 591
Db 592 GCCTTCTGTTTCTCAGCACTGGCATGTTTGTCTCATCATCATGATTCCTCTCTAGTAGC 533
Qy 592 TTCTGTATGTACACTAGCTGTGATAGCCATGACCTGGAGTGTATATGCAAGAACTTCCATT 651
Db 532 TTCTGTATGTACACTAGCTGTGATAGCCATGACCTGGAGTGTATATGCAAGAACTTCCATT 473
Qy 652 GCTGTGTGGAGGATGAGCAGCTGGGCTATCTTAGGCTGGCCATTCAGTGCAGCTCTTGTG 711
Db 472 GCTGTGTGGAGGATGAGCAGCTGGGCTATCTTAGGCTGGCCATTCAGTGCAGCTCTTGTG 413
Qy 712 TTACCCATTGCTTTGATTTGCTGTGTCATGAAACACAGGTGGAGAGTTTCTTTCATTGG 771
Db 412 TTACCCATTGCTTTGATTTGCTGTGTCATGAAACACAGGTGGAGAGTTTCTTTCATTGG 353
Qy 772 TCCTGTATGGCCCTCATATCATTTCTTGTGTCCTGTGTGGTGTATTCAGCTACTATTAT 831
Db 352 TCCTGTATGGCCCTCATATCATTTCTTGTGTCCTGTGTGTGCTCATTCAGCTACTATTAT 293
```

832 GGGAGTTGGTGAATGACCACTCAACATGTTGTTTGTATATATGCTTTACTCTCATGGA 891  
b 292 GGGAGTTGGTGAATGACCACTCAACATGTTGTTTGTATATATGCTTTACTCTCATGGA 233  
y 892 CTGTATCTTTATGTTACAGAACCCCTGGTATTTCTATTTAAATTAATGGAATTTCTGAATTC 951  
b 232 CTGTATCTTTATGTTACAGAACCCCTGGTATTTCTATTTAAATTAATGGAATTTCTGAATTC 173  
y 952 AATGTAGCCCTTTGCTTTGGCTCTCTAGTCTCTAGTCTCTAGTCTCTAGTCTCTATGGAATACCTG 1011  
b 172 AATGTAGCCCTTTGCTTTGGCTCTCTAGTCTCTAGTCTCTAGTCTCTATGGAATACCTG 113  
y 1012 CTGACAGATTTTCATGTTTCAGAAATTTAGCCACCCGTTATTTGGCTTACCTTGGCTCCCAATG 1071  
b 112 CTGACAGATTTTCATGTTTCAGAAATTTAGCCACCCGTTATTTGGCTTACCTTGGCTCCCAATG 53  
y 1072 TATATTTGGTTTATATTTCTTCTATCAGGCTCAAAAGAGGAGATTTCT 1124  
b 52 TATATTTGGTTTAT-ATTTTCTTCTATCAGGCTCAAAAGAGGAGATTTCT 1

EX361838 1080 bp mRNA linear EST 05-MAY-2003  
EX361838 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
Homo sapiens cDNA clone CS0DJ005Y120 5-PRIME, mRNA sequence.  
CESSION  
BX361838  
EX361838.1 GI:30380575  
EST.  
Homo sapiens (human)  
ORCE  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1080)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5955.r for  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DJ005BE10QPI&cluster=5955.r. Contact :  
Feng Liang Email : fliang@life.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DJ005BE10QPI.  
Location/Qualifiers  
1. 1080  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DJ005Y120"  
/cell\_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"  
/cell\_line="JURKAT"  
/clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT  
10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
219 a 238 c 278 g 302 t 43 others

BASE COUNT 219 a 238 c 278 g 302 t 43 others  
ORIGIN  
Query Match 43.4%; Score 850.4; DB 13; Length 1080;  
Best Local Similarity 98.2%; Pred. No. 1.1e-194; Indels 4; Gaps 4;  
Matches 892; Conservative 8; Mismatches 4;  
2 GGCTTGCCATGGCTAGTCGAGGGGCTCGGACGGCTGAAGGGCGGCGGACGCTG 61  
|||||

Db 46 GGGATGCCATGGCTAGTCGAGGGGCTCGGACGGCCTCTGAA - GGSAGCGGGCGCAGCAGTG 104  
Qy 62 GGGATAGCGCCCGGCTGCGGACAACTGCGGAGCTGCTGGGACGCGAGAGGGCG - GC 120  
Db 105 GGGATAGCGCCCGGCTGCGGACAACTGCGGAGCTGCTGGGACGCGAGAGGGCGG - GC 164  
Qy 121 GGGCGGAGCAGCAGGACCGGAGTTATCTGGGAAACAAAGCAGGACAACTCTGGGAC - CTTGAA 180  
Db 165 GGGCGGAGCAGCAGGACCGGAGTTATCTGGGAAACAAAGCAGGACAACTCTGGGAC - CTTGAA 224  
Qy 181 GGATCTACTGCTTTCAAGTCTGCTTTCAAGAGGTATGCTGCTCTCTCTGACGAC - C 240  
Db 225 GGATCTACTGCTTTCAAGTCTGCTTTCAAGAGGTATGCTGCTCTCTCTGACGAC - C 284  
Qy 241 ATCTCTGACTGCTGATGAAACATTTCACTACTGGGAGCACAACACTACCTCAT - CTATGGG 300  
Db 285 ATCTCTGACTGCTGATGAAACATTTCACTACTGGGAGCACAACACTACCTCAT - CTATGGG 344  
Qy 301 GAAGGTTTCAGACTTGGGAATATCCGAGCATATGCCATTCGCTCCTATGCTTAC - CTG 360  
Db 345 GAAGGTTTCAGACTTGGGAATATCCGAGCATATGCCATTCGCTCCTATGCTTAC - CTG 404  
Qy 361 TTGCTTTCATGCTGCGCCAGCTGCAATTTTCATGCAAGAAATTTACAAACTAAT - AAGATTCCT 420  
Db 405 TTGCTTTCATGCTGCGCCAGCTGCAATTTTCATGCAAGAAATTTACAAACTA - AAGATTCCT 464  
Qy 421 GTGTTTTCATTTTTCGATGCTTCTGGCTTTTGTGAGCTGTATTTGTGAACTT - TACTTT 480  
Db 465 GTGTTTTCATTTTTCGATGCTTCTGGCTTTTGTGAGCTGTATTTGTGAACT - TACTTT 524  
Qy 481 TACAAGGCTGTGTCAGAAAGTTTGGGTTGCACTGAGTCGAATGATGCTAGC - TTTCTTG 540  
Db 525 TACAAGGCTGTGTCAGAAAGTTTGGGTTGCACTGAGTCGAATGATGCTAGC - TTTCTTG 584  
Qy 541 GTTCTCAGCACTGGCATGTTTGGCTCATCATCAGCATTCTCTCTAGTAGCT - TCTGTATG 600  
Db 585 GTTCTCAGCACTGGCATGTTTGGCTCATCATCAGCATTCTCTCTAGTAGCT - TCTGTATG 644  
Qy 601 TACACTACGTTGATAGCCATGACTGGAATGTTATGGAACAAAGACTTCCAT - TGTGCTGCTG 660  
Db 645 TACACTACGTTGATAGCCATGACTGGAATGTTATGGAACAAAGACTTCCAT - TGTGCTGCTG 704  
Qy 661 GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTAGTCAGCTCTGTGTTTAC - CCAAT 720  
Db 705 GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTAGTCAGCTCTGTGTTTAC - CCAAT 764  
Qy 721 GCCTTTGATTTGCTGGTCAATGAAACACAGGTGGAGAGTTTCTTTTCACT - GCTGCTGATG 780  
Db 765 GCCTTTGATTTGCTGGTCAATGAAACACAGGTGGAGAGTTTCTTTTCACT - GCTGCTGATG 824  
Qy 781 GCCCTCATCTATTTCT - GGTGCTGCTGGTGGTCAATGACAGCTACTATTTAT - GGGAGTT 839  
Db 825 GCCCTCATCTATTTCTGGGTGCTGGTGGTCAATGACAGCTACTATTTAT - GGGAGTT 884  
Qy 840 GGTGATTCACCACTCAACATTTGTTTGTATATGCTTTTACTCTCT - CATGGACCTGATC 898  
Db 885 GGTGATTCACCACTCAACATTTGTTTGTATATGCTTTTACTCTCTCTCT - CATGGACCTGATC 944  
Qy 899 TTTATGCT 906  
Db 945 TTTATGCT 952

RESULT 11  
EX369641  
LOCUS  
DEFINITION  
BX369641 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
cDNA clone CS0DK002YM06 5-PRIME, mRNA sequence.  
ACCESSION  
BX369641  
VERSION  
BX369641.1 GI:30449834  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

FEATURES	source	Location/Qualifiers
1. .933		
/organism="Homo sapiens"		
/mol_type="mRNA"		
/db_xref="taxon:9606"		
/clone="CS0DK002YM06"		
/cell_type="HELA CELLS		
/call_line="HELA"		
/clone_lib="Homo sapiens HELA CELLS		
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
BASE COUNT	211 a 219 c 197 g 304 t	2 others
ORIGIN		
Query Match	41.6%	Score 816.4; DB 13; Length 933;
Best Local Similarity	96.1%	Pred. No. 1.8e-186;
Matches	854;	Conservative 0; Mismatches 33; Indels 2; Gaps 2;
790	CTATTTCTGGTGCCCTGTGGTGCATATGACAGCTACTATTATGGAGAGTTCGGTGATTGCA	849
28	CTATTTCTGGTGCCCTGTGGTGGTCAATGACAGCTACTATTATGGAGAGTTCGGTGATTGCA	87
850	CCACTCAACATTGTTTGTATATGTCCTTTACTCTCATGGACCTGACTCTTTATGGTACA	909
88	CCACTCAACATTGTTTGTATATGTCCTTTACTCTCATGGACCTGACTCTTTATGGTACA	147
910	GAACCCCTGGTATTTCTATTTAAATAATGAGATTTTCGAATTCGAATGTAGCCTTTGCTTTG	969
148	GAACCCCTGGTATTTCTATTTAAATAATGAGATTTTCGAATTTCAATGTAGCCTTTGCTTTG	207
970	GCTCTCTAGTCTTACACACTGACTTCTCTTATGGAAATACCTGCTCAGAGATTTTCATGTT	1029
208	GCTCTCTAGTCTTACACACTGACTTCTCTTATGGAAATACCTGCTCAGAGATTTTCATGTT	267
1030	CAGAAATTTAGGCCACCCGATTTGGCTTACCTTGGCTCCAAATCTATATTTGGTTTATAATT	1089
268	CAGAAATTTAGGCCACCCGATTTGGCTTACCTTGGCTCCAAATCTATATTTGGTTTATAATT	327
1090	TTCTTTCATCCAGCCTCCAAAGAGGAGAGATTTCTTTTCCCTGTGTATCCACTTATATGT	1149
328	TTCTTTCATCCAGCCTCCAAAGAGGAGAGATTTCTTTTCCCTGTGTATCCACTTATATGT	387
1150	CTCTGTGGCGCTGTGGCTCTCTCTGCACTTCAGAAATGTTACACTTTGTGTTTCAACGA	1209
388	CTCTGTGGCGCTGTGGCTCTCTCTGCACTTCAGAAATGTTTACCACTTTGTGTTTCAACGA	447
1210	TATCGCCTGGAGCACTATCTGTGACATCGAATTTGGCTGGCACTTAGGAACCTGCTTCTCTG	1269
448	TATCGCCTGGAGCACTATCTGTGACATCGAATTTGGCTGGCACTTAGGAACCTGCTTCTCTG	507
1270	TTTGGGCTCTTTGTCAATTTTCTGGCTCTGTGGCACTGTTTCAGAGGATATCAAGGGCCCCCTT	1329
508	TTTGGGCTCTTTGTCAATTTTCTGGCTCTGTGGCACTGTTTCAGAGGATATCAAGGGCCCCCTT	567



```
Db 541 TCCCTGTGTATCCACTATATATGCTCTGTGTGGCGCTGTGGCTCTCTCTGCACTTCAGAAAT 600
QY 1187 GTTACCACTTTGTGTTTCAAGCATATCGCTGAGCACTATCTACTGTGACATCGAATTGSC 1246
Db 601 GTTACCACTTTGTGTTTCAAGCATATCGCTGAGCACTATCTACTGTGACATCGAATTGSC 660
QY 1247 TGGCAATTAGGAACCTGTCTTCCTGTTTGGGCTCTTGTCAATTTTCTCGCTCTGTGGCACTGT 1306
Db 661 TGGCAATTAGGAACCTGTCTTCCTGTTTGGGCTCTTGTCAATTTTCTCGCTCTGTGGCACTGT 720
QY 1307 TCAGAGGATATCA-CGGGCCCTTGATTTGTATCCAGAAATTTTACGAAATGCTACAGAC 1365
Db 721 TCAGAGGATATCANCGGGCCCTTGATTTGTATCCAGAAATTTTACGAAATGCTACAGAC 780
QY 1366 CCAACCATCCACACTGTGCCAGAGGAGCACTGTGTGAA---TGCTGTGTGGGAAAGAG 1422
Db 781 CCAACCATCCACACTGTGCCAGAGGAGCACTGTGTGAAATGCTGTGTGGGAAAGAG 840
QY 1423 TGGTATCGAAT 1433
Db 841 TGGTATCGAAT 851

RESULT 14
BM450315 1016 bp mRNA linear EST 05-FEB-2002
LOCUS AGNCOURT_6394296 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5494628
DEFINITION 5', mRNA sequence.
ACCESSION BM450315
VERSION BM450315.1 GI:18499355
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1016)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM12120 row: d column: 21
High quality sequence stop: 621.
Location/Qualifiers
1. 1016
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5494628"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SF0RT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 239 a 242 c 213 g 321 t 1 others
ORIGIN
Query Match 40.9%; Score 802.8; DB 12; Length 1016;
Best Local Similarity 98.9%; Pred. No. 3.6e-183;
Matches 829; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
QY 704 CTCCTGGTTACCCATGCCCTTGTGATTTGCTGTGTCATGAACACAGGTGAAGATTCT 763
```

```
Db 11 CGCTGGTTTACCCATTGCTTTGATTTGCTGGTCATGAAACACAGGTGGAAGATTCT 70
QY 764 TTCAATTTGGTGGCTGATGGCCCTCATACTATTTCTGGTGGCTGTGGTGTGTCATTGACAGCT 823
Db 71 TTCAATTTGGTGGCTGATGGCCCTCATACTATTTCTGGTGGCTGTGGTGTGTCATTGACAGCT 130
QY 824 ACTATTATGGGAAGTTGGTGAATGACCACTCAACATTTGTTTGTATATATGTCCTTCTC 883
Db 131 ACTATTATGGGAAGTTGGTGAATGACCACTCAACATTTGTTTGTATATATATCTTCTTCTC 190
QY 884 CTCAATGACCTGATCTTTATGTTGACAGAACCTGGTATTTCTTATTTATTAATGATGATTC 943
Db 191 CTCAATGACCTGATCTTTATGTTGACAGAACCTGGTATTTCTTATTTATTAATGATGATTC 250
QY 944 TGAATTTCAATGTAGCTTTGCTTTGGCTCTCTCTAGTCTCTACCACTGACTTCTCTTATGG 1003
Db 251 TGAATTTCAATGTAGCTTTGCTTTGGCTCTCTCTAGTCTCTACCACTGACTTCTCTTATGG 310
QY 1004 AATACCTGCTGCAGAGATTTCAATTTAGCAATTTAGGCCACCCGTTATGGCTTACCTTGG 1063
Db 311 AATACCTGCTGCAGAGATTTCAATTTAGCAATTTAGGCCACCCGTTATGGCTTACCTTGG 370
QY 1064 CTCAATGTATATTTGGTTTATATATTTCTTCTCATCCAGCTTCAAAAGAGGAGATTC 1123
Db 371 CTCAATGTATATTTGGTTTATATATTTCTTCTCATCCAGCTTCAAAAGAGGAGATTC 430
QY 1124 TTTTCCCTGTGTATCCACTATATGTTCTGTGGGCTGTGGCTCTCTCTGCACTTCAGA 1183
Db 431 TTTTCCCTGTGTATCCACTATATGTTCTGTGGGCTGTGGCTCTCTCTGCACTTCAGA 490
QY 1184 AATGTTACCACTTTGTTGTTTCAACGATATCGCTGGAGCACTATCTGTGACATCGAAT 1243
Db 491 AATGTTACCACTTTGTTGTTTCAACGATATCGCTGGAGCACTATCTGTGACATCGAAT 550
QY 1244 GGCTGGATTTAGAACTGCTTCTGTTTGGGCTCTTGTGCACTTTCTGCTGTGGCAC 1303
Db 551 GGCTGGATTTAGAACTGCTTCTGTTTGGGCTCTTGTGCACTTTCTGCTGTGGCAC 610
QY 1304 TGTTCAAGATATACAGGCGCCCTTGATTTGTATCCAGAAATTTACCGAATTTGCTACAG 1363
Db 611 TGTTCAAGATATACAGGCGCCCTTGATTTGTATCCAGAAATTTACCGAATTTGCTACAG 670
QY 1364 ACCCAACATCCACACTGTGCCAGAGGAGGAGCACTGTGAAATGCTGTGGGAAAGAGT 1423
Db 671 ACCCAACATCCACACTGTGCCAGAGGAGGAGCACTGTGAAATGCTGTGGGAAAGAGT 730
QY 1424 GGTAATGATTTCCAGCAGCTTCTTCTTCTCCGACAAATGGCAGCTTCAGTTCAATCCAT 1483
Db 731 GGTAATGATTTCCAGCAGCTTCTTCTTCTCCGACAAATGGCAGCTTCAGTTCAATCCAT 790
QY 1484 -CAGAGTTCAGAGTCAAGTTACC-AAACCTTTTGAGAGGAGCACTTGGCCACCCGG 1539
Db 791 CCAAGATTCAGAGTCAAGTTACC-AAACCTTTTGAGAGGAGCACTTGGCCACCCGG 848

RESULT 15
BM452994
LOCUS AGNCOURT_7888267 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6141418
DEFINITION 5', mRNA sequence.
ACCESSION BM452994
VERSION BM452994.1 GI:22706978
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
```



Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM13461 row: j column: 11  
 High quality sequence stop: 725.  
 Location/Qualifiers  
 1..856  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6141418"  
 /tissue\_type="retinoblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_67"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
 Average insert size 1.75 kb. Library constructed by Life  
 Technologies."  
 FEATURES  
 source 184 a 188 c 189 g 295 t

Query Match 40.4%; Score 791.6; DB 13; Length 856;  
 Best Local Similarity 98.1%; Pred. No. 1.7e-180;  
 Matches 822; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

y	587	GTAGCTTCGTATGTACACTAGCTTGTAGCCATGACCTGGATGGTATATGACCAAGACTT	646
b	1	GTAGCTTCGTATGTACACTAGCTTGTAGCCATGACCTGGATGGTATATGACCAAGACTT	60
y	647	CCATTCCTGTGTGGAGTAGCAGCTGGGGCTATCTTAGCTGGCCATTCAGTCGACCTC	706
b	61	CCATTCCTGTGTGGAGTAGCAGCTGGGGCTATCTTAGCTGGCCATTCAGTCGACCTC	120
y	707	TTGGTTTACCATTCCTTGTATTTGCTGCTCATCAAAACACAGGTGGAAGATTCTTTC	766
b	121	TTGGTTTACCATTCCTTGTATTTGCTGCTCATCAAAACACAGGTGGAAGATTCTTTC	180
y	767	ATTGGTCGTGATGGCCCTCATATTTCTGGTGCCTGTGGTGTCTATTCACAGCTACT	826
b	181	ATTGGTCGTGATGGCCCTCATATTTCTGGTGCCTGTGGTGTCTATTCACAGCTACT	240
y	827	ATTATGGGAAGTTGGTGGATGGACCACTCAACATGTTTGTATATGCTTTTACTCTC	886
b	241	ATTATGGGAAGTTGGTGGATGGACCACTCAACATGTTTGTATATGCTTTTACTCTC	300
y	887	ATGGACCTGATCTTTATGGTACAGAACCTGGTATTTCTATTTAATTAATGATTTCTGA	946
b	301	ATGGACCTGATCTTTATGGTACAGAACCTGGTATTTCTATTTAATTAATGATTTCTGA	360
y	947	ATTTCATGTAGCCTTTGGCTCTCTAGTCTACCACTGACCTCTCTTATGGAT	1006
b	361	ATTTCATGTAGCCTTTGGCTCTCTAGTCTACCACTGACCTCTCTTATGGAT	420
y	1007	ACCTGTGTGACAGATTTCATGTTTCAAGATTAGGCCACCCGGTATTGGCTTACCTTGGCTC	1066
b	421	ACCTGTGTGACAGATTTCATGTTTCAAGATTAGGCCACCCGGTATTGGCTTACCTTGGCTC	480
y	1067	CAATGTATATTGGTTTATATTTTCTTCATCCAGCTCACAAGAGGAGAGATTCTTT	1126
b	481	CAATGTATATTGGTTTATATTTTCTTCATCCAGCTCACAAGAGGAGAGATTCTTT	540
y	1127	TCCTGTGTATCCACTATATGTTCTGTGGGGCTGTGGCTCTCTCTGCACCTTCAGAAAT	1186
b	541	TCCTGTGTATCCACTATATGTTCTGTGGGGCTGTGGCTCTCTCTGCACCTTCAGAAAT	600
y	1187	GTTACCACTTTGTTTCAACGATATCGCTGGAGCACTATATCTGTGACATCGAATTGGC	1246
b	601	GTTACCACTTTGTTTCAACGATATCGCTGGAGCACTATATCTGTGACATCGAATTGGC	660

QY	1247	TGCATTAGGAACCTGCTCTCTCTGTTGGGCTCTTGTCAATTTCTCGCTCTGCGACTGT	1306
Db	661	TGCATTAGGAACCTGCTCTCTCTGTTGGGCTCTTGTCAATTTCTCGCTCTGCGACTGT	720
QY	1307	TCAGAGGATATCAGGGG-CGCCCTTGATTTGATATCCAGAA-TTTTACCGAATTGCTACAGA	1364
Db	721	TCAGAGGATATCAGGGGCCCCCTTGATTTGATATCCAGAAATTTTACCGAATTGCTACAGA	780
QY	1365	CCCAACCATCCACACTGCTCCAGAGGACACCTGTGAATGTCTGTGTGGGAAAGAG	1422
Db	781	ACCAACCATCCACACTGCTCCAGAGGACACCTGTGAATGTCTGTGTGGGAAAG	838

Search completed: January 13, 2004, 06:32:00  
 Job time : 4216 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	58.8	3.0	7218	1	US-08-230-463-14
2	37.8	1.9	4403765	3	US-09-103-840A-2
3	37.8	1.9	4411529	3	US-09-103-840A-1
4	34.8	1.8	801	2	US-08-770-379-16
5	34.8	1.8	801	3	US-08-757-669A-16
6	34.8	1.8	801	4	US-09-298-568-3
7	34.8	1.8	801	4	US-09-230-371A-16
8	34.8	1.8	884	2	US-08-901-200A-11
9	34.8	1.8	884	3	US-09-219-391-11
10	34.8	1.8	6418	1	US-08-480-528A-11
11	34.8	1.8	6418	1	US-08-478-666-11
12	34.8	1.8	6418	5	PCF-US93-10520-11
13	34.6	1.8	5117	3	US-08-854-585-1
14	34.6	1.8	5117	4	US-09-447-533-1
15	34.6	1.8	5117	5	PCF-US95-05512-1
16	34.2	1.7	305	4	US-09-313-294A-7029
17	34.2	1.7	1338	4	US-09-461-325-120
18	33.8	1.7	1377	4	US-09-252-911A-11074
19	33.8	1.7	1407	4	US-09-252-911A-11004
20	33.6	1.7	600	4	US-09-252-991A-5737
21	33.6	1.7	1053	4	US-09-252-991A-1697
22	33.6	1.7	1152	4	US-09-252-991A-1478
23	33.6	1.7	1827	4	US-09-252-991A-5756
24	33.6	1.7	2136	4	US-09-252-91A-5688
25	33.6	1.7	2907	2	US-09-018-628-17
26	33.6	1.7	2907	3	US-09-273-378-17
27	33.6	1.7	2907	3	US-09-018-635-26
					Sequence 14, Appl
					Sequence 1, Appl
					Sequence 16, Appl
					Sequence 3, Appl
					Sequence 16, Appl
					Sequence 11, Appl
					Sequence 11, Appl
					Sequence 11, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 120, App
					Sequence 11074, A
					Sequence 11004, A
					Sequence 5737, Ap
					Sequence 1697, Ap
					Sequence 1478, Ap
					Sequence 5756, Ap
					Sequence 5688, Ap
					Sequence 17, Appl
					Sequence 26, Appl
					Sequence 16, Appl

```

RESULT 3
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 441529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Query Match 1.9%; Score 37.8; DB 3; Length 441529;
Best Local Similarity 57.0%; Pred. No. 38;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 11 TGGCTAGTCGAGGGGCTCGGCACGCGCTCAAGGGCAGCGGGCCAGCAGTGGGGATACGG 70
Db 2261400 TGACGGGTGGAGGGGGCGACGCCAGCCAGCGCAGGCGCAACGATCGGGCGTGGGAAACGG 2261341
QY 71 CCCCAGGCTGCGGACCAAGCTCGCGGAGCTCTCTGGGCAGCCGAGAGCGGGCGCGCGGAGC 130
Db 2261340 GTCTGGCAGTCCCGATCTCGGCGACGACGCCGGGCGGCGACGCTCCCGAGCGGGCG 2261281
QY 131 A 131
Db 2261280 A 2261280

RESULT 4
US-08-770-379-16/c
; Sequence 16, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooper & Durham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.

```

REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 801 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
JS-08-770-379-16

Query Match 1.8%; Score 34.8; DB 2; Length 801;  
Best Local Similarity 53.7%; Pred. No. 1.3;  
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
2Y 2 GGCTTGCCATGCTAGTCGAGGGGCTCGGCAGCGCTGGAAGGCGCAGCGGGGCCAGCAGTG 61  
Db 143 GCCCGCGGAGGAGCGCGCGGCGCCCTGCCCCCGCGGGCGGCGGCTCCGGCCGGG 84  
2Y 62 GGGATACGCGCCCGGCTGCGGACAAGCTGCGGAGCTGCTGGGCGAGCCCGGAGAGCGGGCG 121  
Db 83 GCCTGCGGCGAGCGCGCGGAGCGGAGCGGCGGGGAGGGGCGGCGCGCGTGTGGGGG 24  
2Y 122 GCGCGGAGCACCGG 135  
Db 23 GCGCGGGCGCGGG 10

RESULT 5  
JS-08-757-669A-16/c  
Sequence 16, Application US/08757669A  
Patent No. 6183751  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,669A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 801 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
US-08-757-669A-16

Query Match 1.8%; Score 34.8; DB 3; Length 801;  
Best Local Similarity 53.7%; Pred. No. 1.3;  
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 2 GGCTTGCCATGCTAGTCGAGGGGCTGCGGAGCGCTGGAAGGCGCAGCGGGGCCAGCAGTG 61  
Db 143 GCCCGCGGAGGAGCGCGCGGCGCCCTGCCCCCGCGGGCGGCGGCTCCGGCCGGG 84  
QY 62 GGGATACGCGCCCGGCTGCGGACAAGCTGCGGAGCTGCTGGGCGAGCCCGGAGAGCGGGCG 121  
Db 83 GCCTGCGGCGAGCGCGCGGAGCGGAGCGGCGGGGAGGGGCGGCGCGTGTGGGGG 24  
QY 122 GCGCGGAGCACCGG 135  
Db 23 GCGCGGGCGCGGG 10

RESULT 6  
US-09-298-568-3/c  
Sequence 3, Application US/09298568  
Patent No. 6322792  
GENERAL INFORMATION:  
APPLICANT: Kieff, Elliott D.  
APPLICANT: Ballestar, Mary E.  
APPLICANT: Kaye, Kenneth M.  
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO  
FILE REFERENCE: 16412-10001R  
CURRENT APPLICATION NUMBER: US/09/298,568  
EARLIER FILING DATE: 1999-04-21  
EARLIER APPLICATION NUMBER: US 60/109,422  
EARLIER FILING DATE: 1998-11-19  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 3  
LENGTH: 801  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-298-568-3

Query Match 1.8%; Score 34.8; DB 4; Length 801;  
Best Local Similarity 53.7%; Pred. No. 1.3;  
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 2 GGCTTGCCATGCTAGTCGAGGGGCTCGGCAGCGCTGGAAGGCGCAGCGGGGCCAGCAGTG 61  
Db 143 GCCCGCGGAGGAGCGCGCGGCGCCCTGCCCCCGCGGGCGGCGGCTCCGGCCGGG 84  
QY 62 GGGATACGCGCCCGGCTGCGGACAAGCTGCGGAGAGCTGCTGGGCGAGCCCGGAGAGCGGGCG 121  
Db 83 GCCTGCGGCGAGCGCGCGGAGCGGAGCGGCGGGGAGGGGCGGCGGCGGCTGTGGGGG 24  
QY 122 GCGCGGAGCACCGG 135  
Db 23 GCGCGGGCGCGGG 10

RESULT 7  
US-09-230-371A-16/c  
Sequence 16, Application US/09230371A  
Patent No. 6348586  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S

APPLICANT: Moore, Patrick S  
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND  
 TITLE OF INVENTION: USES THEREOF  
 FILE REFERENCE: 45185-G-PCT-US  
 CURRENT APPLICATION NUMBER: US/09/230,371A  
 CURRENT FILING DATE: 1999-11-17  
 PRIOR APPLICATION NUMBER: PCT/US97/13346  
 PRIOR FILING DATE: 1997-07-22  
 NUMBER OF SEQ ID NOS: 30  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 16  
 LENGTH: 801  
 TYPE: DNA

ORGANISM: Kaposi's sarcoma-associated herpesvirus  
 S-09-230-371A-16

Query Match 1.8%; Score 34.8; DB 4; Length 801;  
 Best Local Similarity 53.7%; Pred. No. 1.3;  
 Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
 Y 2 GGCTTCCCATGCTAGTCGAGGGGCTCGGAGCGGCTGAAGGGCAGCGGGGCCAGCAGTG 61  
 b 143 GCCCGCGAGGAGCGCGGCGGCGGCTGCCCCCGGCGGCGGCTCGGCGCGG 84  
 Y 62 GGATACGGCCCGGCTGCGGACAAGCTGCGGAGCTCTGCGGAGCGCGGCGGCGG 121  
 b 83 GCTGCGGCGAGCGCGGCGGAGCGGAGCGGAGCGGAGCGGCGGCGGCGGCGG 24  
 Y 122 GCGCGGAGCACCGG 135  
 b 23 GCGCGGCGGCGG 10

RESULT 8

S-08-901-200A-11/c  
 Sequence 11, Application US/08901200A  
 Patent No. 5854071  
 GENERAL INFORMATION:  
 APPLICANT: OPPERMANN, HERMANN  
 APPLICANT: OZKAYNAK, ENGIN  
 APPLICANT: KUBERASAMPATH, THANGAVEL  
 APPLICANT: RUEGER, DAVID C.  
 APPLICANT: PANG, ROY H.L.  
 APPLICANT: COHEN, CHARLES M.  
 TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
 STREET: 45 SOUTH STREET  
 CITY: HOPKINTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 01748

COMPUTER READABLE FORM: disk  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/901,200A  
 FILING DATE: 28-JUL-1997  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MEYERS, THOMAS C.  
 REGISTRATION NUMBER: 36,989  
 REFERENCE/DOCKET NUMBER: CRP-076DV2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (508) 435-9001  
 TELEFAX: (508) 435-6951  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 884 base pairs  
 TYPE: nucleic acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1..884  
 OTHER INFORMATION: /note= "hop-2 genomic sequence"  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1..837  
 OTHER INFORMATION: /note= "EXON ONE"  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 884  
 OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN  
 OTHER INFORMATION: POSITIONS 884 IN THIS SEQUENCE AND POSITION 1 IN SEQ ID NO 14"  
 US-08-901-200A-11

Query Match 1.8%; Score 34.8; DB 2; Length 884;  
 Best Local Similarity 52.0%; Pred. No. 1.4;  
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
 QY 1 AGGCTTCCCATGCTAGTCGAGGGGCTCGGAGCGGCTGAAGGGCAGCGGGGCCAGCAGT 60  
 Db 500 AGCGGGGGGCTCAGCTGGGGCGCTCAGCGGGCGGCGATCGGCTCCGCGCGGCCCA 441  
 QY 61 GGGGATACGGCCCGGCTGCGGACAAGCTCGGGAGCTGCTGGGAGCGCGGAGCGGGC 120  
 Db 440 GGGCTGGGAGCGCGCGGCGGCAAGAGGCTGGGCTCGCGGGCGGCGGCGGCGGCG 381  
 QY 121 GCGCGGAGCACCGGAGCGGAGTTATCTGGG 150  
 Db 380 GGGCGGAGCGGCGGCGGCGGCTCAGCG 351

RESULT 9

US-09-219-391-11/c  
 Sequence 11, Application US/09219391  
 Patent No. 6153583  
 GENERAL INFORMATION:  
 APPLICANT: OPPERMANN, HERMANN  
 APPLICANT: OZKAYNAK, ENGIN  
 APPLICANT: KUBERASAMPATH, THANGAVEL  
 APPLICANT: RUEGER, DAVID C.  
 APPLICANT: PANG, ROY H.L.  
 APPLICANT: COHEN, CHARLES M.  
 TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
 STREET: 45 SOUTH STREET  
 CITY: HOPKINTON  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 01748  
 COMPUTER READABLE FORM: disk  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/219,391  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/901,200  
 FILING DATE: 28-JUL-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MEYERS, THOMAS C.  
 REGISTRATION NUMBER: 36,989  
 REFERENCE/DOCKET NUMBER: CRP-076DV2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (508) 435-9001

TELEFAX: (508) 435-6951  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 884 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..884  
OTHER INFORMATION: /note= "hOP-2 genomic sequence"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1..837  
OTHER INFORMATION: /note= "EXON ONE"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 884..885  
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN  
OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1088..1277  
OTHER INFORMATION: /note= "EXON TWO"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1350..1814  
OTHER INFORMATION: /note= "EXON THREE"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1834..1835  
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN  
OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1883..2077  
OTHER INFORMATION: /note= "EXON FOUR"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 2902..2981  
OTHER INFORMATION: /note= "EXON FIVE"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3507..3617  
OTHER INFORMATION: /note= "EXON SIX"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 6116..6361  
OTHER INFORMATION: /note= "EXON SEVEN"  
US-08-480-528A-11

Query Match 1.8%; Score 34.8; DB 3; Length 884;  
Best Local Similarity 52.0%; Pred. No. 1.4;  
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 AGGCTTGCGATGGCTAGTCGAGGGGCTCGGCGAGCGCTGAAGGCGAGCGGGGCGCAGT 60  
DB 500 AGCGCGGGGGCTCAGCTGGGGGCTCAGCGGGCGGCATCGGCTCCGGCGCGACCCA 441

QY 61 GGGGATACGGCCCCGGCTGCGGACAGCTGCGGGAGCTGTGGGCGACCCGAGAGCGGGC 120  
DB 440 GGGCTTGGGGAGCGCCCCGACGCGGAGGAGGCTGGGCTCGGCGGGCGGGCGGGC 381

QY 121 GCGCGGAGCACCAGCGACCGAGTTATCTGGG 150  
DB 380 GGGCGGAGACGGGCGACCGCGGCTCAGCG 351

RESULT 10  
US-08-480-528A-11/c  
Sequence 11, Application US/08480528A  
Patent No. 5652118  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMANN  
APPLICANT: OZKAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: PANG, ROY H. L.  
APPLICANT: COHEN, CHARLES M.  
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
STREET: 45 SOUTH STREET  
CITY: HOPKINTON  
STATE: MA  
COUNTRY: USA  
ZIP: 01748  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,528A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FENTON Esq., GILLIAN M.  
REGISTRATION NUMBER: 36,508  
REFERENCE/DOCKET NUMBER: CRP-076FW

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7560  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6418 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..6361  
OTHER INFORMATION: /note= "hOP-2 genomic sequence"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1..837  
OTHER INFORMATION: /note= "EXON ONE"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 884..885  
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN  
OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1088..1277  
OTHER INFORMATION: /note= "EXON TWO"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1350..1814  
OTHER INFORMATION: /note= "EXON THREE"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1834..1835  
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN  
OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1883..2077  
OTHER INFORMATION: /note= "EXON FOUR"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 2902..2981  
OTHER INFORMATION: /note= "EXON FIVE"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3507..3617  
OTHER INFORMATION: /note= "EXON SIX"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 6116..6361  
OTHER INFORMATION: /note= "EXON SEVEN"  
US-08-480-528A-11

Query Match 1.8%; Score 34.8; DB 1; Length 6418;  
Best Local Similarity 52.0%; Pred. No. 5.4;  
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 AGGCTTGCGATGGCTAGTCGAGGGGCTCGGCGAGCGCTGAAGGCGAGCGGGGCGCAGT 60  
DB 500 AGCGCGGGGGCTCAGCTGGGGGCTCAGCGGGCGGCATCGGCTCCGGCGCGACCCA 441

QY 61 GGGGATACGGCCCCGGCTGCGGACAGCTGCGGGAGCTGTGGGCGAGCCGAGAGCGGGC 120  
DB 440 GGGCTTGGGGAGCGCCCCGACGCGGAGGAGGCTGGGCTCGGCGGGCGGGCGGGC 381

QY 121 GCGCGGAGCACCAGCGACCGAGTTATCTGGG 150  
DB 380 GGGCGGAGACGGGCGACCGCGGCTCAGCG 351

RESULT 11  
US-08-479-666-11/c

```

LOCATION: 2902..2981
FEATURE:
NAME/KEY: exon
LOCATION: 3507...3617
OTHER INFORMATION: /note= "EXON SIX"
FEATURE:
NAME/KEY: exon
LOCATION: 6116..6361
OTHER INFORMATION: /note= "EXON SEVEN"
US-08-4793-666-11

Query Match 1.8%; Score 34.8; DB 1; Length 6418;
Best Local Similarity 52.0%; Pred. No. 5.4;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 AGGCTTGCCATGGCTAGTGCAGGGGCTCGGCAGCGCCTGAGGGGCGAGGGGGGCCAGCAGT 60
DB 500 AGGCGGGGGCGCTACGCTGGGCGCTCAGCGGGCGCGCATCGGCTCCGCGCGGACCCA 441
QY 61 GGGGATACGGCCCCGGTGGCGGACAGCTGCGGAGCTCTGCGGACGCCGAGAGCGGGC 120
DB 440 GGGCGTGGGGACGCCCGACGGCGGCAAGCGCTGGGCTCGCGGGCGGGCGGGCGGGC 381
QY 121 GCGCGGGAGCACCGGACCGCAGTTATCTGGG 150
DB 380 GGGGCGGACGGGCGACCGCGGCTCAGCG 351

RESULT 12
PCT-US93-10520-11/c
Sequence 11, Application PC/TUS9310520
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: CREATIVE BIOMOLECULES, INC.
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10520
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/667,274
FILING DATE: 11-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,764
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/753,059
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,857
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/923,780
FILING DATE: 31-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,813
FILING DATE: 31-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829

```



REFERENCE/DOCKET NUMBER: CRP-0762C

TELECOMMUNICATION INFORMATION:

TELEPHONE: 4508/435-9001

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 6418 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..6361

OTHER INFORMATION: /note= "HOP-2 genomic sequence"

FEATURE:

NAME/KEY: exon

LOCATION: 1..837

OTHER INFORMATION: /note= "EXON ONE"

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 884..1885

OTHER INFORMATION: /note= "A Gap Occurs Between

OTHER INFORMATION: Positions 884 and 885 in this Sequence"

FEATURE:

NAME/KEY: exon

LOCATION: 1088..1277

OTHER INFORMATION: /note= "EXON TWO"

FEATURE:

NAME/KEY: exon

LOCATION: 1350..1814

OTHER INFORMATION: /note= "EXON THREE"

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1834..1835

OTHER INFORMATION: /note= "A Gap Occurs Between

OTHER INFORMATION: Positions 1834 and 1835 in this Sequence"

FEATURE:

NAME/KEY: exon

LOCATION: 1883..2077

OTHER INFORMATION: /note= "EXON FOUR"

FEATURE:

NAME/KEY: exon

LOCATION: 2902..2981

OTHER INFORMATION: /note= "EXON FIVE"

FEATURE:

NAME/KEY: exon

LOCATION: 3507..3617

OTHER INFORMATION: /note= "EXON SIX"

FEATURE:

NAME/KEY: exon

LOCATION: 6116..6361

OTHER INFORMATION: /note= "EXON SEVEN"

PCT-US93-10520-11

Query Match 1.8%; Score 34.9; DB 5; Length 6418;

Best Local Similarity 52.0%; Pred. No. 5.4; Mismatches 0; Gaps 0;

Matches 78; Conservative 0; Indels 72; Gaps 0;

Qy 1 AGGCTTGCCATGGCTAGTCGAGGGGCTCGGCAGCGCCTGAAGGGGCGAGCGGGCCAGCACT 60

Db 500 AGCGCGGGGGCGCTCAGCTGGGGCGCTCAGCGGGCGGCATCGGCTCGCGCGCGACCA 441

Qy 61 GGGGATACGGCCCGCGCTGCGGACAGCTGCGGAGCTGCTGGGAGCCGAGAGCGGGC 120

Db 440 GGGCTGGGGAGCGCCCGACCGAGGAGGCTGGGCTCGCGCGGCGCGCGGGGCG 381

Qy 121 GCGCGGAGCAGCCGAGCCAGCTATCTGGG 150

Db 380 GGGGGGAGCGGGCGACCGCGGCTCAGCG 351

RESULT 13

US-08-854-585-1

; Sequence 1, Application US/08854585

; Patent No. 6114140

; GENERAL INFORMATION:

; APPLICANT: Tonks, Nicholas K. and stman, Arne

; TITLE OF INVENTION: Density Enhanced Protein Tyrosine Phosphatase

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 233 South Wacker Drive, Suite 6300

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/854,585

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/237,940

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Borun, Michael F.

; REGISTRATION NUMBER: 25,447

; REFERENCE/DOCKET NUMBER: 27866/31954

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-474-6300

; TELEFAX: 312-474-0448

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5117 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 350..4364

; US-08-854-585-1

Query Match 1.8%; Score 34.6; DB 3; Length 5117;

Best Local Similarity 54.3%; Pred. No. 5.4;

Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 30 GCAGCGCTGAAGGGCAGCGGGCGGCGGATAGTGGGATACGGCCCGGCTGGCGGCAAGCT 89

Db 123 GCGGCCCCCGCGGCGGCGGCTAGGCTCGGCGTGTGGCGGCGCGCGCGCGCT 182

Qy 90 GCGGAGCTGCTGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 149

Db 183 GCCATGTCTCCGGGAAGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 242

Qy 150 GAACAAGC 158

Db 243 GGAGAGGC 251

RESULT 14

US-09-447-533-1

; Sequence 1, Application US/09447533

; Patent No. 6552169

; GENERAL INFORMATION:

; APPLICANT: Tonks, Nicholas K.

; Ostman, Arne

; TITLE OF INVENTION: DENSITY ENHANCED PROTEIN TYROSINE

; PHOSPHATASES

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed IP Law Group PLLC

; STREET: Suite 6300, 701 Fifth Avenue



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM nucleic - nucleic search, using sw model

Run on: January 14, 2004, 13:55:05 ; Search time 658 Seconds  
(without alignments)  
10504.652 Million cell updates/sec

Title: US-09-922-225A-1  
Perfect score: 1961  
Sequence: 1 agccttgcatggtagtcg.....aggcttttgatgataaaaa 1961

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1961	100.0	1961	11	US-09-922-225A-1
2	1017	51.9	1017	11	US-09-922-225A-7
3	935.4	47.7	1225	11	US-09-922-225A-5
4	897.6	45.8	953	15	US-10-050-704-35
5	580.4	29.6	880	11	US-09-922-225A-3
6	444.2	22.7	530	11	US-09-918-995-22803
7	371	18.9	372	9	US-09-777-564-39
8	371	18.9	372	15	US-10-015-219-39
9	291	14.8	445	15	US-10-102-521-980
10	237	12.1	500	13	US-10-029-386-10276
11	122.4	6.2	124	13	US-10-029-386-10276
12	90	4.6	3317	15	US-10-128-714-173
13	90	4.6	4016	15	US-10-128-714-173
14	60	3.1	60	13	US-09-908-975-8397
15	57.4	2.9	427	11	US-09-918-995-37352

16	50.6	2.6	51	11	US-09-922-225A-101	Sequence 101, App
17	50.6	2.6	51	11	US-09-922-225A-102	Sequence 102, App
18	50.6	2.6	51	11	US-09-922-225A-103	Sequence 103, App
19	43.2	2.2	5856	13	US-10-311-455-789	Sequence 789, App
20	42.2	2.2	671	15	US-10-184-644-346	Sequence 346, App
21	42.2	2.2	671	15	US-10-184-644-346	Sequence 346, App
22	41.2	2.1	1089	15	US-10-184-644-266	Sequence 266, App
23	41.2	2.1	1089	15	US-10-184-644-266	Sequence 266, App
24	40.8	2.1	25554	12	US-10-017-161-695	Sequence 695, App
25	40.8	2.1	84510	12	US-10-292-798-607	Sequence 607, App
26	39	2.0	1041	13	US-10-027-632-121553	Sequence 121553,
27	39	2.0	1041	13	US-10-027-632-121553	Sequence 121553,
28	38.8	2.0	1254	13	US-10-027-632-204193	Sequence 204193,
29	38.8	2.0	1254	13	US-10-027-632-204193	Sequence 204193,
30	38	1.9	775	13	US-10-140-472-120	Sequence 120, App
31	38	1.9	775	13	US-10-141-761-120	Sequence 120, App
32	38	1.9	775	13	US-10-142-885-120	Sequence 120, App
33	38	1.9	775	13	US-10-158-790-120	Sequence 120, App
34	38	1.9	775	13	US-10-137-871-120	Sequence 120, App
35	38	1.9	775	13	US-10-140-805-120	Sequence 120, App
36	38	1.9	775	13	US-10-140-864-120	Sequence 120, App
37	38	1.9	775	13	US-10-140-923-120	Sequence 120, App
38	38	1.9	775	13	US-10-141-756-120	Sequence 120, App
39	38	1.9	775	13	US-10-141-759-120	Sequence 120, App
40	38	1.9	775	13	US-10-123-155-120	Sequence 120, App
41	38	1.9	775	16	US-10-146-731-120	Sequence 120, App
42	37.8	1.9	18301	12	US-10-242-355-708	Sequence 708, App
43	37.4	1.9	426	15	US-10-156-761-190	Sequence 190, App
44	37.4	1.9	7006	13	US-10-341-434-221	Sequence 221, App
45	37.4	1.9	8733	13	US-10-144-198-11	Sequence 11, App

ALIGNMENTS

RESULT 1  
US-09-922-225A-1  
; Sequence 1, Application US/09922225A  
; Publication No. US20030104385A1  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
; TITLE OF INVENTION: Associated with Bipolar Disorder  
; FILE REFERENCE: P-EA 4572  
; CURRENT APPLICATION NUMBER: US/09/922,225A  
; CURRENT FILING DATE: 2003-01-14  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1961  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10)...(1842)  
US-09-922-225A-1

Query Match		100.0%	Score 1961;	DB 11;	Length 1961;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 1961;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	1	AGGCTTGCCATGGCTAGTCGAGGGGCTCGGAGCGCCCTGAAGGGCAGCGGGGCCAGCAGT	60		
Db	1	AGGCTTGCCATGGCTAGTCGAGGGGCTCGGAGCGCCCTGAAGGGCAGCGGGGCCAGCAGT	60		
Qy	61	GGGATACGGCCCGGCTCGGACAGCTCGGGAGCTGCTGGGAGCGGAGCGGGCC	120		
Db	61	GGGATACGGCCCGGCTCGGACAGCTGCTGGGAGCTGCTGGGAGCGGAGCGGGCC	120		
Qy	121	GGCGCGGAGCACCAGCGCCGAGTATCTGGGAACAAAGCAGGCAAGTCTGGGCACCTGAA	180		
Db	121	GGCGCGGAGCACCAGCGCCGAGTATCTGGGAACAAAGCAGGCAAGTCTGGGCACCTGAA	180		

181	GGATCTACTGCTTTTCAAGTGTCTGCTTTTTCAGCAAGGTTATGTGCTGCTCTCTCTGAGCAAC	240
181	GGATCTACTGCTTTTCAAGTGTCTGCTTTTTCAGCAAGGTTATGTGCTGCTCTCTCTGAGCAAC	240
241	ATCTCTGACGTGTGATGAAACATTCACACTCTGGAGCCCAACACACTACCTCATCTATGCG	300
241	ATCTCTGACGTGTGATGAAACATTCACACTCTGGAGCCCAACACACTACCTCATCTATGCG	300
301	GAAGGGTTTTCAGACTTGGGAATATTCGCCAGCATATGCCAATTCGGCTCTCTATGCTTACCTG	360
301	GAAGGGTTTTCAGACTTGGGAATATTCGCCAGCATATGCCAATTCGGCTCTCTATGCTTACCTG	360
361	TTGCTTTCATGCTGGCCAGCTGCATTCATGCAAGAAATCTCAACAACTAATGAAGATCTT	420
361	TTGCTTTCATGCTGGCCAGCTGCATTCATGCAAGAAATCTCAACAACTAATGAAGATCTT	420
421	GTGTTTTTACTTTTTTGGCATGCTTCTTGGCTTTTTTGTGAGCTGTATTTTGAACCTTTACCTT	480
421	GTGTTTTTACTTTTTTGGCATGCTTCTTGGCTTTTTTGTGAGCTGTATTTTGAACCTTTACCTT	480
481	TACAAGGCTGTGTGCAAGAAATTTGGGTTTGCACTGAGTGGAAATGATGATGCTTACCTTCTG	540
481	TACAAGGCTGTGTGCAAGAAATTTGGGTTTGCACTGAGTGGAAATGATGATGCTTACCTTCTG	540
541	GTTCTTCAGCACTGGCATGTTTTTGGCTCATCATCAGCAATTCCTTCTAGTAGTCTTCTGTATG	600
541	GTTCTTCAGCACTGGCATGTTTTTGGCTCATCATCAGCAATTCCTTCTAGTAGTCTTCTGTATG	600
601	TACACTAGCTTGAATAGCAATGACATGGAATGATATGGAACAAGACTTCATCTGCTGTGCTG	660
601	TACACTAGCTTGAATAGCAATGACATGGAATGATATGGAACAAGACTTCATCTGCTGTGCTG	660
661	GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTCAGTGCAGCTCTTGTTGTTACCCAT	720
661	GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTCAGTGCAGCTCTTGTTGTTACCCAT	720
721	GCTTTTGAATTTGCTGGTCAAGAAACAGAGTGGAGAGTTTCTTTCATGGTCCGTGATG	780
721	GCTTTTGAATTTGCTGGTCAAGAAACAGAGTGGAGAGTTTCTTTCATGGTCCGTGATG	780
781	GCCCTCATACTATTTCTGGTCCCTGCTGGTGCATTGCACAGCTACTATATGGAAGTTG	840
781	GCCCTCATACTATTTCTGGTCCCTGCTGGTGCATTGCACAGCTACTATATGGAAGTTG	840
841	GTGATGACCACTCAACATTTGTTTGTATAATGCTTTACTCCTCATGCAACCTGATCTT	900
841	GTGATGACCACTCAACATTTGTTTGTATAATGCTTTACTCCTCATGCAACCTGATCTT	900
901	TATGTTACAGAACCTTGATTTCTATTAATTAATGGAATTCCTGAAATTTCAATGTAGCC	960
901	TATGTTACAGAACCTTGATTTCTATTAATTAATGGAATTCCTGAAATTTCAATGTAGCC	960
961	TTTGTCTTTGGCTCTCTTAGTCTACCACTGACTTCTTTATGGAATACTGCTGCAAGAGA	1020
961	TTTGTCTTTGGCTCTCTTAGTCTACCACTGACTTCTTTATGGAATACTGCTGCAAGAGA	1020
1021	TTTTCATGTTTCAAGAAATTTAGGCCACCCGTTATGGCTTACCTTGGCTCCAAATGATATTTGG	1080
1021	TTTTCATGTTTCAAGAAATTTAGGCCACCCGTTATGGCTTACCTTGGCTCCAAATGATATTTGG	1080
1081	TTTATAAATTTTCTTCAATCCAGCTCACAAAGAGGAGATTTCTTTTCCCTGTGTATPCCA	1140
1081	TTTATAAATTTTCTTCAATCCAGCTCACAAAGAGGAGATTTCTTTTCCCTGTGTATPCCA	1140
1141	CTTATATGTTCTGTGGGCTGTGGCTCTCTCTGCACTTCAGAAATGTTTACCACTTTGTG	1200
1141	CTTATATGTTCTGTGGGCTGTGGCTCTCTCTGCACTTCAGAAATGTTTACCACTTTGTG	1200
1201	TTTTCACGATATCGGCTGAGCACTATAGTGTGACATCGAATTCGGCTGGCAATTAGGAAC	1260
1201	TTTTCACGATATCGGCTGAGCACTATAGTGTGACATCGAATTCGGCTGGCAATTAGGAAC	1260
1261	GTCTTCTGCTTTTGGGCTCTTGTTCATTTTTTCTCGCTGTGGCACTGTTTCAGAGGATATCAC	1320

Db	1361	GTCTTCCTGTTGGGCTCTTGTCATTTTTCGCTCTGTGGCACTGTTTCAGAGNATATCAC	1320
Qy	1321	GGGCCCCCTTGATTTGTATCCAGAAATTTACCGAATTTGCTACAGACCCACCATCCACACT	1380
Db	1321	GGGCCCCCTTGATTTGTATCCAGAAATTTACCGAATTTGCTACAGACCCACCATCCACACT	1380
Qy	1381	GTCCAGAAAGGCGAGACTGTGAATGTCGTGTGGGAAAAGAGTGGTATCGATTTCCGACG	1440
Db	1381	GTCCAGAAAGGCGAGACTGTGAATGTCGTGTGGGAAAAGAGTGGTATCGATTTCCGACG	1440
Qy	1441	AGCTTCCTTTCTTCTGTGACAAATGGCGAGCTTCAGTTCATTTCCATCAGAGTTCAGAGTCAAG	1500
Db	1441	AGCTTCCTTTCTTCTGTGACAAATGGCGAGCTTCAGTTCATTTCCATCAGAGTTCAGAGTCAAG	1500
Qy	1501	TTACCAAAAACCTTTTGCAGAAAGACACTCTGGCCACCCGGATTTGTTCTACTGACATGAAT	1560
Db	1501	TTACCAAAAACCTTTTGCAGAAAGACCTCTGGCCACCCGGATTTGTTCTACTGACATGAAT	1560
Qy	1561	GACCAGAACTTAGAAGAGCCATCCAGATATATTTGATATATGATTAATGCCATTTATTAGTG	1620
Db	1561	GACCAGAACTTAGAAGAGCCATCCAGATATATTTGATATATGATTAATGCCATTTATTAGTG	1620
Qy	1621	GATTTGGACACCATGACAGAGAAAACACCCCGGGAGCCAAAATATTCATCCATAAAGAAGAA	1680
Db	1621	GATTTGGACACCATGAGAGAAAACACCCCGGGAGCCAAAATATTCATCCATAAAGAAGAA	1680
Qy	1681	TGGATACAGCTTGGCCCTATAGACCAATTCCTTTGATGCTTCTAGATCTTCAAAAGCTGCTGGCG	1740
Db	1681	TGGATACAGCTTGGCCCTATAGACCAATTCCTTTGATGCTTCTAGATCTTCAAAAGCTGCTGGCG	1740
Qy	1741	GCATTCATGTGCCCCCTTCTGTGCAGATCAGTATACAGTGTACGTAAACTACACCATCCTC	1800
Db	1741	GCATTCATGTGCCCCCTTCTGTGCAGATCAGTATACAGTGTACGTAAACTACACCATCCTC	1800
Qy	1801	AAACCCCGGAAAGCAAAACCAATCAGGAAAGAAAAGTGGAGGTTAGCAACACACCTGTGGC	1860
Db	1801	AAACCCCGGAAAGCAAAACCAATCAGGAAAGAAAAGTGGAGGTTAGCAACACACCTGTGGC	1860
Qy	1861	CCCAAGGACAAACCATCTTGTTTAACTATTGATTCAGCTGACCTGACCTCCCTGCAAGTCAT	1920
Db	1861	CCCAAGGACAAACCATCTTGTTTAACTATTGATTCAGCTGACCTGACCTCCCTGCAAGTCAT	1920
Qy	1921	GGCCTGTAACTTTGTAAATAAAGGTCTTCTGACATGA AAAA	1961
Db	1921	GGCCTGTAACTTTGTAAATAAAGGTCTTCTGACATGA AAAA	1961

```

RESULT 2
US-09-922-225A-7
; Sequence 7, Application US/09922225A
; Publication NO. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; TITLE OF INVENTION: Associated with Bipolar Disorder
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1017)
US-09-922-225A-7

```

Query Match 51.9%; Score 1017; DB 11; Length 1017;  
Best Local Similarity 100.0%; Pred. No. 4.9e-303;  
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



b 1001 GGGGGCAATCTATGTCCTTCTGTCAGATCAGTATACAGTGTACGTAAACTACACCAT 1060  
y 1797 CTTCAAAACCCGGAAGCAAAAGCAATCAGGAAGAGAAAGTGGAGTTAGCAACACACCTG 1856  
b 1061 CTTCAAAACCCGGAAGCAAAAGCAATCAGGAAGAGAAAGTGGAGTTAGCAACACACCTG 1120  
y 1857 TGGCCCAAGAACCAACCATCTTGTAACTATTTGATTCAGTGACCTGACTCCCTGCAAG 1916  
b 1121 TGGCCCAAGAACCAACCATCTTGTAACTATTTGATTCAGTGACCTGACTCCCTGCAAG 1180  
y 1917 TCATGCGCTGTAACTTTGTATAAAGGTCTTCGACATGAAAA 1961  
b 1181 TCATGCGCTGTAACTTTGTATAAAGGTCTTCGACATGAAAA 1225

RESULT 4

S-10-050-704-35  
Sequence 35, Application US/10050704  
Publication No. US20030050442A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
FILE OF INVENTION: 62 Human Secreted Proteins  
PCT REFERENCE: P2039P1  
CURRENT APPLICATION NUMBER: US/10/050,704  
CURRENT FILING DATE: 2002-01-18  
PRIOR APPLICATION NUMBER: 09/684,524  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: PCT/US00/08979  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/128,693  
PRIOR FILING DATE: 1999-04-09  
PRIOR APPLICATION NUMBER: 60/130,991  
PRIOR FILING DATE: 1999-04-26  
NUMBER OF SEQ ID NOS: 344  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 953  
TYPE: DNA  
ORGANISM: Homo sapiens  
S-10-050-704-35

Query Match 45.8%; Score 897.6; DB 15; Length 953;  
Best Local Similarity 99.9%; Pred. No. 3.8e-266;  
Matches 897; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
y 1064 CTCCAATGATATATTTGGTTTAAATTTCTTCATCCAGCCTCACAAAGAGAGATTTC 1123  
b 1 CTCCAATGATATATTTGGTTTAAATTTCTTCATCCAGCCTCACAAAGAGAGATTTC 60  
y 1124 TTTTCCCTGTATCCACTATATGCTCTGTGGCGCTGTGCTCTCTCTGACTTCAGA 1183  
b 61 TTTTCCCTGTATCCACTATATGCTCTGTGGCGCTGTGCTCTCTCTGACTTCAGA 120  
y 1184 AATGTTACCACCTTTGTGTTTAAACGATATCGCTGGAGCACTATATCTGTGACATG 1243  
b 121 AATGTTACCACCTTTGTGTTTAAACGATATCGCTGGAGCACTATATCTGTGACATG 180  
y 1244 GGTGCGCATTTAGAACTGCTCTCTGTTGGCTCTTGTCTATTTCTGCTCTGTGSCAC 1303  
b 181 GGTGCGCATTTAGAACTGCTCTCTGTTGGCTCTTGTCTATTTCTGCTCTGTGGAC 240  
y 1304 TGTTCAGAGATATACGGGGCCCTTTGATTTGTATCCAGAAATTTTACCGAAATTCACAG 1363  
b 241 TGTTCAGAGATATACGGGGCCCTTTGATTTGTATCCAGAAATTTTACCGAAATTCACAG 300  
y 1364 ACCCAACCATCCACATGTCGCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1423  
b 301 ACCCAACCATCCACATGTCGCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
y 1424 GGTATCCGATTTCCAGAGAGCTTCTCTCTTCCAGAAATTTGGAGGCTTCAGTTTCATTCAT 1483  
b 361 GGTATCCGATTTCCAGAGAGCTTCTCTCTTCCAGAAATTTGGAGGCTTCAGTTTCATTCAT 420

QY 1484 CAGAGTTCAGAGGTCAAGTTACCAAAACCTTTTGCAGAGAGAGCTCTGGCCACCCGGATTG 1543  
Db 421 CAGAGTTCAGAGGTCAAGTTACCAAAACCTTTTGCAGAGAGAGCTCTGGCCACCCGGATTG 480  
QY 1544 TTCCTACTGACATGAATGACAGAACTTAGAGAGCCATCCAGATATATTGATATCAGTA 1603  
Db 481 TTCCTACTGACATGAATGACAGAACTTAGAGAGCCATCCAGATATATTGATATCAGTA 540  
QY 1604 AATGCCATTATTTAGTGGATTGGACACCATGAGAGAAACACCCCGGAGCCAAATATT 1663  
Db 541 AATGCCATTATTTAGTGGATTGGACACCATGAGAGAAACACCCCGGAGCCAAATATT 600  
QY 1664 CATCAATAAAGAGAGATGATCAGCTTGGCCCTATAGACCATTCCTTTGATGCTTCAGAT 1723  
Db 601 CATCAATAAAGAGAGATGATCAGCTTGGCCCTATAGACCATTCCTTTGATGCTTCAGAT 660  
QY 1724 CTTCAAAGCTGCTCGGGCAATTTATGTCCTTCTGTCAGATCAGTATACAGTGTACG 1783  
Db 661 CTTCAAAGCTGCTCGGGCAATTTATGTCCTTCTGTCAGATCAGTATACAGTGTACG 720  
QY 1784 TAAACTACCATCTCTCAAAACCCCGGAAAGCAAAATCAGGAAGAAAGTGGAGGTT 1843  
Db 721 TAAACTACCATCTCTCAAAACCCCGGAAAGCAAAATCAGGAAGAAAGTGGAGGTT 780  
QY 1844 AGCAACACACCTGTGGCCCAAGGACCAACCTTTGTTAACTATTGATTCAGTCACT 1903  
Db 781 AGCAACACACCTGTGGCCCAAGGACCAACCTTTGTTAACTATTGATTCAGTCACT 840  
QY 1904 GACTCCCTGCAAGTCATGCGCTGTAACTTTGTAATAAGGTCTCTGACATGAAAA 1961  
Db 841 GACTCCCTGCAAGTCATGCGCTGTAACTTTGTAATAAGGTCTCTGACATGAAAA 898

RESULT 5

US-09-922-225A-3  
Sequence 3, Application US/09922225A  
Publication No. US20030104385A1  
GENERAL INFORMATION:  
APPLICANT: Evans, Glen A.  
FILE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
TITLE OF INVENTION: Associated with Bipolar Disorder  
FILE REFERENCE: P-EA 4672  
CURRENT APPLICATION NUMBER: US/09/922,225A  
CURRENT FILING DATE: 2003-01-14  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 880  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (113)...(877)  
US-09-922-225A-3

Query Match 29.6%; Score 580.4; DB 11; Length 880;  
Best Local Similarity 98.8%; Pred. No. 4.3e-168;  
Matches 595; Conservative 0; Mismatches 6; Indels 1; Gaps 1;  
QY 1017 GAGATTTCATGTTCAAGATTTAGGCCACCCGGTATTTGGCTTACCTTGGCTCCAAATGTATAT 1076  
Db 280 GTGTGTGCCAGTTCAAGATTTAGGCCACCCGGTATTTGGCTTACCTTGGCTCCAAATGTATAT 339  
QY 1077 TTGGTTTATAATTTTCTTCATCCAGCCTCACAAAGAGAGAGATTTCTTTTCCCTGTGTA 1136  
Db 340 TTGGTTTATAATTTTCTTCATCCAGCCTCACAAAGAGAGAGATTTCTTTTCCCTGTGTA 399  
QY 1137 TCACCTTATATGCTCTGTGGCGCTGGCTCTCTCTGCACCTTCAGAAATGTTACCACTT 1196  
Db 400 TCACCTTATATGCTCTGTGGCGCTGGCTCTCTCTGCACCTTCAGAAATGTTACCACTT 459  
QY 1197 TGTGTTTCAACGATATCGCTTGGAGCACTATCTGTGACATCGAATTTGGCTGGCATTAGG 1256

460 TGTGTTCAACGATATCGCCTGGAGCACTATATCTGTGACATCGAATTTGGCTGGCATTAGG 519  
1257 AACTGCTTCTCTGTTGGGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTCAGAGGATA 1316  
520 AACTGCTTCTCTGTTGGGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTCAGAGGATA 579  
1317 TCAGGCGCCCTTGATTTGTATCCAGAAATTTACCGAATTTGTCAGACCCCAACCATCCA 1376  
580 TCAGGCGCCCTTGATTTGTATCCAGAAATTTACCGAATTTGTCAGACCCCAACCATCCA 639  
1377 CACTGTCCAGAGGAGCACTCTGATCTGTGTGGGAAAGAGTGTATCGAATTTCC 1436  
640 CACTGTCCAGAGGAGCACTCTGATCTGTGTGGGAAAGAGTGTATCGAATTTCC 699  
1437 CAGCAGCTTCTCTTCTCTGACAAATTTGGCAGCTTCAGTTCATTTCCATCAGAGTTCCAGAGG 1496  
700 CAGCAGCTTCTCTTCTCTGACAAATTTGGCAGCTTCAGTTCATTTCCATCAGAGTTCCAGAGG 759  
1497 TCAGTTACCAAAACCTTTTCAGAGGAGCTCTGGCCAGCCGGATTTCTTCTACTGACAT 1556  
760 TCAGTTACCAAAACCTTTTCAGAGGAGCTCTGGCCAGCCGGATTTCTTCTACTGACAT 819  
1557 GAATGACCAAGATCTAGAGAGCCATCCAGATATATTGATATCAGTAAATGCCATTATTT 1616  
820 GAATGACCAAGATC-AGAAGAGCCATCCAGATATATTGATATCAGTAAATGCCATTATTT 878  
1617 AG 1618  
879 AG 880  
RESULT 6  
US-09-918-995-22803  
; Sequence 22803, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-755  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 22803  
; LENGTH: 530  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(530)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-22803  
Query Match 22.7%; Score 444.2; DB 11; Length 530;  
Best Local Similarity 95.0%; Pred. No. 3.8e-126;  
Matches 477; Conservative 0; Mismatches 4; Indels 21; Gaps 1;  
2Y 1083 TATAATTTTCTTCATCCAGCTCACAAGAGGAGAGATTTCTTTTCCCTGTGTATCCACT 1142  
28 TATAATTTTCTTCATCCAGCTCACAAGAGGAGAGATTTCTTTTCCCTGTGTATCCACT 87  
1143 TATATGCTCTGTGGCGCTGTGGCTCTCTCTGCACTTCAGACAGTTTCTGTACTTCCA 1181  
88 TATATGCTCTGTGGCGCTGTGGCTCTCTCTGCACTTCAGACAGTTTCTGTACTTCCA 147  
1182 GAATGTTTACCACCTTTGTGTTTCAACGATATCGCGTGGAGCACTATATCTGTGACATCGAA 1241  
148 GAATGTTTACCACCTTTGTGTTTCAACGATATCGCGTGGAGCACTATATCTGTGACATCGAA 207  
1242 TTGGCTGGCAATTAGGAACGTGCTTCCTGTTTGGGCTCTTGTGCAATTTTCTCGCTCTGTGGC 1301

208 TTGGCTGGCAATTAGGAACGTGCTTCCTGTTTGGGCTCTTGTCAATTTCTCGCTCTGTGGC 267  
1302 ACTGTTTCAGAGATATCAGGGCCCTTTGATTGTATCCAGAAATTTACCGAATTTGCTTAC 1361  
268 ACTGTTTCAGAGATATCAGGGCCCTTTGATTGTATCCAGAAATTTACCGAATTTGCTTAC 327  
1362 AGACCCAAACCATCCACACTGTCCAGAAAGGAGAGACTGTGAATGTCTGTGTGGGAAAGA 1421  
328 AGACCCAAACCATCCACACTGTCCAGAAAGGAGAGACTGTGAATGTCTGTGTGGGAAAGA 387  
1422 GTGGTATCCATTTCCAGCAGCTTCTTCTTCCAGCAATTTGGCAGCTTCAGTTCATTTCC 1481  
388 GTGGTATCCATTTCCAGCAGCTTCTTCTTCCAGCAATTTGGCAGCTTCAGTTCATTTCC 447  
1482 ATCAGAGTTTCAGAGGTTCAGTTACCAAAACCTTTTSCAGAGGAGCTCTGGCCACCCGGAT 1541  
448 ATCAGAGTTTCAGAGGTTCAGTTACCAAAACCTTTTSCAGAGGAGCTCTGGCCACCCGGGA 507  
1542 TGTTCTCTACTGACATGAATGAC 1563  
508 GGTTCTCTACTGACATGAATGAC 529  
RESULT 7  
US-09-777-564-39/c  
; Sequence 39, Application US/09777564  
; Patent No. US2002022591A1  
; GENERAL INFORMATION:  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.493  
; CURRENT APPLICATION NUMBER: US/09/777,564  
; CURRENT FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 1730  
; SOFTWARE: FastSeq for Window Version 4.0  
; SEQ ID NO 39  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(372)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-777-564-39  
Query Match 18.9%; Score 371; DB 9; Length 372;  
Best Local Similarity 99.7%; Pred. No. 1.3e-103;  
Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1532 CCACCCGGATTGTTCTCTACTGACATGAATGACCAATCTAGAAGAGCCATCCAGATATA 1591  
372 CCACCCGGATTGTTCTCTANTGACATGAATGACCAATCTAGAAGAGCCATCCAGATATA 313  
1592 TTGATATCAGTAATGCCATTTATTTAGTGGATTGGACACCATGAGAGAAACACCCCGG 1651  
312 TTGATATCAGTAATGCCATTTATTTAGTGGATTGGACACCATGAGAGAAACACCCCGG 253  
1652 AGCCAAATATTCATCCAATAAAGAGAAATGGATCAGCTTGGCCTATAGACCATTCCTTG 1711  
252 AGCCAAATATTCATCCAATAAAGAGAAATGGATCAGCTTGGCCTATAGACCATTCCTTG 193  
1712 ATGCTTCTAGATCTTCAAGAGCTGTGGGGCTTTCTATGTCCTCCCTTCTGTCAGATCAGT 1771  
192 ATGCTTCTAGATCTTCAAGAGCTGTGGGGCTTTCTATGTCCTCCCTTCTGTCAGATCAGT 133  
1772 ATACAGTGTACGTAATAACTACCATCTCTCAAAACCCCGGAAAGCAAAATCAGGAAGA 1831  
132 ATACAGTGTACGTAATAACTACCATCTCTCAAAACCCCGGAAAGCAAAATCAGGAAGA 73  
1832 AAGTGGAGGTAGCAACACACCTGTGGGCCCAAGGACAAACCACTTTGTTAACTATTGA 1891



b 72 AAAGTGAGGTTAGACACACACCTGTGCCCCAAGAGCACCACTCTGTACTATTGA 13  
y 1892 TTCCAGTGACCT 1903  
b 12 TTCCAGTGACCT 1

RESULT 8  
S-10-015-219-39/c  
Sequence 39, Application US/10015219  
Publication No. US2003008299A1  
GENERAL INFORMATION:  
APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.493C1  
CURRENT APPLICATION NUMBER: US/10/015,219  
CURRENT FILING DATE: 2002-03-02  
NUMBER OF SEQ ID NOS: 1739  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 39  
LENGTH: 372  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 354  
OTHER INFORMATION: n = A,T,C or G  
S-10-015-219-39

Query Match 18.9%; Score 371; DB 15; Length 372;  
Best Local Similarity 99.7%; Pred. No. 1.3e-103;  
Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1532 CCACCCGGATTGTCCTACTGACATGAATGACCAATCTAGAAGAGCCATCCAGATATA 1591  
b 372 CCACCCGGATTGTCCTANTGACATGAATGACCAATCTAGAAGAGCCATCCAGATATA 313  
y 1592 TTGATATCAGTAATGCCATTATTAGTGGATTGGACACCATGAGAAACACCCCGGG 1651  
b 312 TTGATATCAGTAATGCCATTATTAGTGGATTGGACACCATGAGAAACACCCCGGG 253  
y 1652 AGCAAAATATTCATCCATAAAGAGATGGATCAGCTTGGCCCTATAGACCATTCCTTG 1711  
b 252 AGCAAAATATTCATCCATAAAGAGATGGATCAGCTTGGCCCTATAGACCATTCCTTG 193  
y 1712 ATGCTTCTAGATCTTCAAGCTGCTGGGGCAATCTATGTCCCTTCTGTGCATCAGT 1771  
b 192 ATGCTTCTAGATCTTCAAGCTGCTGGGGCAATCTATGTCCCTTCTGTGCATCAGT 133  
y 1772 ATACAGTGTAGCTTAACTACACCATCTCAAAACCCCGGAAAGCAAGCAATCAGGAAGA 1831  
b 132 ATACAGTGTAGCTTAACTACACCATCTCAAAACCCCGGAAAGCAAGCAATCAGGAAGA 73  
y 1832 AAAGTGAGGTTAGCAACACACCTGTGGCCCCCAAGGCAACCATCTGTGTTAACTATTGA 1891  
b 72 AAAGTGAGGTTAGCAACACACCTGTGGCCCCCAAGGCAACCATCTGTGTTAACTATTGA 13

RESULT 9  
S-10-102-524-980  
Sequence 980, Application US/10102524  
Publication No. US20030109434A1  
GENERAL INFORMATION:  
APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.493C1  
CURRENT APPLICATION NUMBER: US/10/015,219  
CURRENT FILING DATE: 2002-03-02  
NUMBER OF SEQ ID NOS: 1739  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 39  
LENGTH: 372  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 354  
OTHER INFORMATION: n = A,T,C or G  
S-10-015-219-39

Query Match 18.9%; Score 371; DB 15; Length 372;  
Best Local Similarity 99.7%; Pred. No. 1.3e-103;  
Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

APPLICANT: Harlocker, Susan L.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
FILE REFERENCE: 210121.572  
CURRENT APPLICATION NUMBER: US/10/102,524  
CURRENT FILING DATE: 2002-03-19  
NUMBER OF SEQ ID NOS: 1863  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 980  
LENGTH: 445  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-102-524-980

Query Match 14.8%; Score 291; DB 15; Length 445;  
Best Local Similarity 100.0%; Pred. No. 7.9e-79;  
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 CTGCATTTTCATGCAAGAAATCTACAAACTAATAGATCTCTGTGTTTACTTTTCGGAT 439  
DB 155 CTGCATTTTCATGCAAGAAATCTACAAACTAATAGATCTCTGTGTTTACTTTTCGGAT 214  
QY 440 GTCTTCTGGCTTTTGTGAGCTGTATTTGTGAATTTTACTTTTACAAGGCTGTGCAAGA 499  
DB 215 GTCCTCTGGCTTTTGTGAGCTGTATTTGTGAATTTTACTTTTACAAGGCTGTGCAAGA 274  
QY 500 AGTTTGGTTTCAGTGAGTGAATGATGCTAGCTTCTGTGTTTCTCAGCACTGGCATGT 559  
DB 275 AGTTTGGTTTCAGTGAGTGAATGATGCTAGCTTCTGTGTTTCTCAGCACTGGCATGT 334  
QY 560 TTGCTCATCATCAGCATTCCTTCTAGTAGCTTCTGTATGTACACTAGCTTGTATAGCCA 619  
DB 335 TTGCTCATCATCAGCATTCCTTCTAGTAGCTTCTGTATGTACACTAGCTTGTATAGCCA 394  
QY 620 TGACTGGATGTATATGACCAAGACTTCCATTTGCTGTGCTGGGAGTAGCAG 670  
DB 395 TGACTGGATGTATATGACCAAGACTTCCATTTGCTGTGCTGGGAGTAGCAG 445

RESULT 10  
US-10-029-386-10276/c  
Sequence 10276, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David K.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
FILE REFERENCE: AEMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 10276  
LENGTH: 500  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC000381.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.48  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7  
OTHER INFORMATION: NT HIT: gi13376061, EVALUATE 1.00e-127  
OTHER INFORMATION: SWISSPROT HIT: Q86677, EVALUATE 2.10e+00  
OTHER INFORMATION: EST\_HUMAN HIT: AA886018.1, EVALUATE 0.00e+00  
US-10-029-386-10276

Query Match 12.1%; Score 237; DB 13; Length 500;  
Best Local Similarity 98.0%; Pred. No. 4.2e-62;  
Matches 240; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1716 TTCTAGATCTTCAAGCTGTGCGGGGATTTATGTCCTTCTGTGATCAGATCAGTATAC 1775





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 05:21:49 ; Search time 18 seconds

(without alignments)  
1596.295 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGSSGDT.....VNTILKPRKAKQIRKISGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1064	32.5	603	YTH3 CAEEL	P54002 caenorhabdi
2	973	23.7	577	YFV5 SCHPO	Q9P7Q9 schizosacch
3	736	22.5	555	ALG9 YEAST	P53868 saccharomyc
4	214.5	6.6	616	YOC2 YEAST	P30777 saccharomyc
5	144.5	4.4	516	SNP3 YEAST	Q04174 saccharomyc
6	128.5	3.9	551	YN81 YEAST	P53730 saccharomyc
7	123.5	3.8	533	YADC SCHPO	Q09837 schizosacch
8	116	3.5	662	YCOB BUCAL	P57543 buchnera ap
9	115	3.5	386	BSY3 ACEXY	Q9WX70 acetobacter
10	114	3.5	692	SEA9 HUMAN	P48067 homo sapien
11	113.5	3.5	469	NUON BUCAL	P57264 buchnera ap
12	112	3.4	590	NU5M TRYBB	P04540 trypanosoma
13	111	3.4	460	YAGM ECOLI	P75683 escherichia
14	110.5	3.4	496	Y048 UREPA	Q9PR97 ureaplasma
15	110	3.4	633	SEA9 MOUSE	P28571 mus musculu
16	109.5	3.3	672	ASG8 BAT	P58428 rattus norv
17	109.5	3.3	695	FSHR HUMAN	P23945 homo sapien
18	109	3.3	314	O1Q1 HUMAN	Q15612 homo sapien
19	108.5	3.3	544	ALG6 YEAST	Q12001 saccharomyc
20	108	3.3	408	YLO1 SCHPO	Q13883 schizosacch
21	108	3.3	633	SEA9 RAT	P28572 rattus norv
22	108	3.3	669	NU5M MARPO	P26849 marchantia
23	107.5	3.3	501	NUOM BUCAP	Q8K946 buchnera ap
24	107.5	3.3	638	SEA9 BOVIN	Q28039 bos taurus
25	107.5	3.3	1465	YH85 SCHPO	Q9P5N0 schizosacch
26	106.5	3.3	426	YX35 SCHPO	Q09712 schizosacch
27	106.5	3.3	502	PUPP SALTY	P10502 salmonella
28	106	3.2	387	SRB3 BRARE	Q91918 brachydanio
29	106	3.2	692	NU5C MARPO	P06264 marchantia
30	105.5	3.2	692	FSHR RAT	P20395 rattus norv
31	105	3.2	433	TCR STRAU	P02983 staphylococ
32	104.5	3.2	368	CYB TOXGO	Q00672 toxoplasma
33	104.5	3.2	458	TCR STRPN	P11063 streptococc

ALIGNMENTS

RESULT 1

YTH3 CAEEL STANDARD; PRT; 603 AA.  
ID YTH3 CAEEL  
AC P54002;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Putative glycosyl transferase Cl4A4.3 in chromosome II (EC 2.-.-.-).  
GN Cl4A4.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA McMurray A.A.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Durbin R.;  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
CC reticulum (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 22.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z49909; CAA90107.2; -;  
CC PIR; T19245; T19245.  
CC WormPep; C14A4.3; CB32796.  
CC Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane;  
CC Endoplasmic reticulum  
FT TRANSMEM 109 129 POTENTIAL.  
FT TRANSMEM 131 151 POTENTIAL.  
FT TRANSMEM 167 187 POTENTIAL.  
FT TRANSMEM 196 216 POTENTIAL.  
FT TRANSMEM 246 266 POTENTIAL.  
FT TRANSMEM 311 331 POTENTIAL.  
FT TRANSMEM 344 364 POTENTIAL.  
FT TRANSMEM 371 391 POTENTIAL.  
FT TRANSMEM 398 418 POTENTIAL.  
SQ SEQUENCE 603 AA; 68912 MW; 6AFDD9F995CEOC5 CRC64;

Query Match 32.5%; Score 1064; DB 1; Length 603;  
Best Local Similarity 39.4%; Pred. No. 6.5e-68;  
Matches 226; Conservative 96; Mismatches 212; Indels 40; Gaps 10;

QY 48 NKACQVAPGSGTAFKCLLSARLCALLSNISCDFTFYWEPHYLYLYGEGFOTWEVSP 107



412 LETDNNVT-----NVCVGKWTYRPTFFLPDNRSLKFKVSEPDGILPGCFVESNST 464  
 508 -----ATRIPTDNDONLEPSRYIDISCHYLVDL--DTMRETPREPKYSSNKEEWS 560  
 465 WNRREGYQIPEHNEPNEPTRYSLSCDPLIDLEFDHSHKATNVEPIYSKS-DGSHIP 523  
 561 LAYPPFDASRSSKLLRAFYVPLPSQYTYVNYTILKPKK 601  
 524 VMVYFFIDTKQTPMGRAFAVPIEPKMGY-BILVKKPVK 563

RESULT 3

ALG9\_YEAST  
 ID ALG9\_YEAST STANDARD; PRT; 555 AA.  
 AC P53868;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable mannose 6-phosphate transferase ALG9 (EC 2.4.1.-).  
 EN ALG9 OR YNL219C OR N1295.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 XX NCBI\_TaxID=4932;  
 XX [1]  
 XX SEQUENCE FROM N.A.  
 XX STRAIN=SS328.  
 XX MEDLINE=96293493; PubMed=8692962;  
 RA Burda P., Te Heesen S., Brachat A., Wach A., Duesterhoeft A., Abbi M.;  
 RT "Stepwise assembly of the lipid-linked oligosaccharide in the  
 RT endoplasmic reticulum of Saccharomyces cerevisiae: identification of  
 RT the ALG9 gene encoding a putative mannose 6-phosphate transferase";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:7160-7165(1996).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,  
 RA Hilbert H., Moestl D.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CATALYZES THE TRANSFER OF MANNOSE FROM DOL-P-MAN TO  
 CC LIPID-LINKED OLIGOSACCHARIDES.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 22.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X96417; CRA65277.1; --  
 DR EMBL; 271495; CRA96122.1; --  
 DR PIR; S63177; S63177.  
 DR SGD; S0005163; ALG9.  
 DR GO; GO:0005783; C:endoplasmic reticulum; IMP.  
 DR GO; GO:0000030; F:mannosyltransferase activity; IMP.  
 DR GO; GO:0006486; P:protein amino acid glycosylation; IMP.  
 KW Transferase; Glycosyltransferase; Transmembrane;  
 Endoplasmic reticulum.  
 FT TRANSMEM 8 28 POTENTIAL.  
 FT TRANSMEM 63 83 POTENTIAL.  
 FT TRANSMEM 87 107 POTENTIAL.  
 FT TRANSMEM 114 134 POTENTIAL.  
 FT TRANSMEM 177 197 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 269 289 POTENTIAL.  
 FT TRANSMEM 317 337 POTENTIAL.  
 FT TRANSMEM 350 370 POTENTIAL.  
 FT TRANSMEM 555 AA; 63776 MW; 108ED4E0B0C2AAA7 CRC64;  
 SQ SEQUENCE

Query Match 22.5%; Score 736; DB 1; Length 555;  
 Best Local Similarity 33.1%; Pred. No. 9e-45;  
 Matches 179; Conservative 82; Mismatches 199; Indels 80; Gaps 15;  
 QY 76 SNLSDCDEFTNWEPTHYLYGEGFQWTEYSPAVAIRSVAYVILLHAWPAAFHARLQTNK 135  
 DB 26 SLISDCDEFTNWEPLNLVRGFGQTWEYSPFYSIRSWAFLL-----PFYCLYIPVK 79  
 QY 136 IL-----VFYFLRCLLAFVSCICELYFYKAVCKFKGLHVSRRMMLAFVLVSTGMFCSSA 189  
 DB 80 FTDLSEHWNFFITRACLGPFSPFIMEFKLHRETAGSLAQIANIWIIFQLPFGWPHASVE 139  
 QY 190 FLSSSECMYTTLIAMTGMKDTSTAVLGAAG-----AAILGWPSAALG 234  
 DB 140 LLSAVAMLL-----YVGATRSLRYLSTGSTNSTKSLAYNFLASILGWPFVLLS 191  
 QY 235 LPTAFDLLVMKRWKSWFFWSLMALILFLV---PYVVIDSYYYGKLVIAPLNIVLYNVT 291  
 DB 192 LPCLHYL-FNHRILITRTAFDCLIFSLTAFATVITDSIFYGKLAPVSNWILFYNVIN 250  
 QY 292 ---PHGPDLYGTPEWYFYLINGFINNVAFALALVLPLTSMELVLLQRFHVQNLGHPY 348  
 DB 251 ASSESGNIFGVEPWYYPPLNLLNF-----PLPLVLVLAIGIFHUR-----LW 294  
 QY 349 LTLAPYIWIIFIIFFIOPHKERFLFPVYVPLILCGAVALSALQKCYHFVQRYRLEHYTV 408  
 DB 295 PLWASLFTWIAVFTQOPHKERFLYPIYGLITLSASIAFYKVLNLFN-----RKPI 345  
 QY 409 TSNWALGTVFFGLLSPRSVALFRGHHGDLIDYEPYRIATDTTHIVPEGRVNVTV 468  
 DB 346 LKXGKILSVLLIVAGAMGRIVALLVNNYTPATVAYEQFSSLNQGGVKAPV-----VNYCT 400  
 QY 469 GKWEYFPSPSFLPDNWLQFIPSEFRGQLPKFAE-GPLATRI--VPTDMNDQNLSEPS 525  
 DB 401 GREWHFSPSFLPDNHLKFKVSGFDGLPGDFESSGIFPKIETLPKGMNKNIVDTG 460  
 QY 526 RYDIDSKHYLDLDTMRTPRE---PKYSSNKEWISLAVRPPDASRSSKLLRAFYVP 582  
 DB 461 KEWPIRCDFIDIVAPINLTKDVFNPLHL--MDWNKLACAAFDGNSKILGRAFYVP 518

RESULT 4

YGO2\_YEAST  
 ID YGO2\_YEAST STANDARD; PRT; 616 AA.  
 AC P30777;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 72.6 kDa protein in MRFL-HUL5 intergenic region.  
 GN YGL142C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 XX NCBI\_TaxID=4932;  
 XX [1]  
 XX SEQUENCE FROM N.A.  
 XX STRAIN=S288C / FY1769;  
 XX MEDLINE=97197983; PubMed=9046099;  
 RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
 RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
 RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,  
 RT TIF1, MRP1 genes and six new open reading frames";  
 RL Yeast 13:177-182(1997).  
 RL [2]  
 RP SEQUENCE OF 339-616 FROM N.A.  
 RP MEDLINE=9117110; PubMed=1475194;  
 RA Pel H.J., Maat M.J., Rep M., Grivell L.A.;  
 RT "The yeast nuclear gene MRP1 encodes a mitochondrial peptide chain  
 RT release factor and cures several mitochondrial RNA splicing  
 RT defects";  
 RL Nucleic Acids Res. 20:6339-6346(1992).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: SOME, TO S.POMBE SPAC2G11.09, SPAC4G8.12C AND

SPPC16A11.06C.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X99960; CAA68220.1; -;  
EMBL; 272664; CAA96854.1; -;  
EMBL; X60381; CAA42931.1; -;  
PIR; S64156; S64156.  
SGD; S0003110; GP110.  
GO; GO:0008506; P:GPI anchor biosynthesis; IMP.  
InterPro; IPR005599; PMP.  
Pfam; PF03901; PMP; 1.  
Hypothetical protein; Transmembrane.  
TRANSMEM 17 37 POTENTIAL.  
TRANSMEM 97 107 POTENTIAL.  
TRANSMEM 137 157 POTENTIAL.  
TRANSMEM 189 209 POTENTIAL.  
TRANSMEM 241 261 POTENTIAL.  
TRANSMEM 279 299 POTENTIAL.  
TRANSMEM 339 359 POTENTIAL.  
TRANSMEM 393 413 POTENTIAL.  
TRANSMEM 424 444 POTENTIAL.  
CONFLICT 339 339 S -> A (IN REF. 2).  
SEQUENCE 616 AA; 72565 MW; 141ABF6A01B008ED CRC64;

Query Match 6.6%; Score 214.5; DB 1; Length 516;  
Best Local Similarity 22.5%; Pred. No. 7.1e-08;  
Matches 145; Conservative 79; Mismatches 242; Indels 179; Gaps 31;

62 PKCLLSARLCALLS-NISDCDETNYEPHYLYIGEGFOTWEYSPAYAIRSVAY---- 116  
18 FWVFLAFLVNLAVLRTFQADFEQALPEAHKAFKYGELTWEK--FGVRSILFFMIF 75  
117 -----LLH-----AWPAAFHARILQTNKIL----- 137  
76 ELTYRLVLSILLHAYALLSTIGSDLLILLPKVELSQVAEDKRLPFDVTRSPYY 135  
138 -VFYFRLCLLAVSCICELFYKAVCKKFGHLVSR-----MMLAFVLVS 190  
136 GVIYAPKIVNAVLASIGEYIVRFQKLYLTLDRKNEEERRSRGLSEITKALLLSL 195  
181 TGMF-C--SSSAFLPSFCVYTTLIAM-----TGWMYDKTS--IAVLGVAAGAILGWFFS 230  
196 TNFFNCFITRTFI-NSPEMILTSIALYWDWTGGQMIKESSTKSLIFAPLACLQRPSS 254  
231 AAL-----GLPIAFDLLVMKRWKSPHSLMALILFLV--PWVIDSYKGLVIAPINI 284  
255 GLIWIPIPSLILNLVGRKQYHLLFITTSKVLRSFVLFTANALIDMYFEKVTFFPFRF 314  
285 VLYNVETPGPDLYGEPWPVYFLNGF---LNFNV-APALALLVLPLTSLMEYLLQRFHV 340  
315 LKENFTPLS-KFYGVAPWHFFQSLPVLGASIPAFAGLF-PPLS-----KRSFP 365  
341 QNIGHYWLTLAPMYIWFIIFFIQPKERFLFPVPLICGVALSALOKCHVFQOR 400  
366 KKYLNPFQFQKLLINLLNVSTLPHKEPFFIPLQPLFILSSFGLLRLDRDY----- 419  
401 YRLHVTVTSNMLAL-GTVLFGLLSF-SRSVALFRGVHGLDLYPEYRIATDPTIHTV 458  
420 -----WKELSGLSLLWLPVFSVFIAL-----LDTTHSGSIEVWKFLHEE 462  
459 PGRPNVVCVKEWYRPP-----SSFL-----LPDNWLOQFIPSEFQGLKPAFAGPLA 508  
463 PE-----IDSLGFTMPCHSTPGQSYLRSIDQLWSITCNP-----PLHLGLDPEA 508  
509 TRIVPTDMNDQNEEPSRYIDISKHYLDLDMTRTPREP--KYSSNKEWIS----- 560

Db 509 YSKLETYMD-----ESDHLYYDISAFIYKNFPFPPFRKDLSPGKTYKYS---HEWPTYLVVFE 561  
Qy 561 -----LAYRPFL-----DASRSKLLRAFYVPF 583  
Db 562 HMEAFKFLDKSSYIEYRFFNSLAHWDSRRSGDIIYYKLPF 606

RESULT 5

SMP3\_YEAST STANDARD; PRT; 516 AA.  
AC Q04174; Q99400;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SMP3 protein.  
GN SMP3 OR YOR149C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1] SEQUENCE FROM N.A.  
RP STRAIN=NEW5;  
RC MEDLINE=9117125; PubMed=2005867;  
RA Irie K., Araki H., Oshima Y.;  
RT "Mutations in a Saccharomyces cerevisiae host showing increased  
RT holding stability of the heterologous plasmid pSR1.";  
RL Mol. Gen. Genet. 225:257-265(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / FY1678;  
RA Ayadi A., Bordonne R., Camasses A., Madania A., Poch O.,  
RA Tarassov I.A., Winsor B., Martin R.P.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ESSENTIAL PROTEIN INVOLVED IN PLASMID MAINTENANCE WITH  
CC SMP2.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: TO S.POMBE SPAC48.12C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X58121; CAA41123.1; -;  
EMBL; U55020; AAC49635.1; -;  
EMBL; 275057; CAA99355.1; -;  
PIR; S67037; S67037.  
SGD; S0005675; SMP3.  
GO; GO:0006276; P:plasmid maintenance; IMP.  
InterPro; IPR005599; PMP.  
Pfam; PF03901; PMP; 1.  
KW Transmembrane.  
FT TRANSMEM 6 26 POTENTIAL.  
FT TRANSMEM 61 81 POTENTIAL.  
FT TRANSMEM 176 196 POTENTIAL.  
FT TRANSMEM 211 231 POTENTIAL.  
FT TRANSMEM 271 291 POTENTIAL.  
FT TRANSMEM 296 316 POTENTIAL.  
FT TRANSMEM 318 338 POTENTIAL.  
FT TRANSMEM 349 369 POTENTIAL.  
FT CONFLICT 122 123 MQ -> IK (IN REF. 1).  
FT CONFLICT 163 163 E -> G (IN REF. 1).  
FT CONFLICT 169 169 S -> R (IN REF. 1).  
FT CONFLICT 279 279 V -> L (IN REF. 1).  
SQ SEQUENCE 516 AA; 59900 MW; 8D8404622CB69534 CRC64;

Query Match 4.4%; Score 144.5; DB 1; Length 516;  
Best Local Similarity 20.7%; Pred. No. 0.0052;  
Matches 104; Conservative 71; Mismatches 176; Indels 151; Gaps 27;





Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
A Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
A James K., Jones L., Mungall K., Murphy L., McDonald S., McLean J.,  
A Mooney P., Moule S., Mungall K., Murphy L., McDonald S., McLean J.,  
A Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
A Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
A Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
A Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
A Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
A Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
A Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
A Borzym K., Langer I., Beck A., Lehnach R., Reinhardt R., Pohl T.M.,  
A Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
A Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
A Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
A Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
A Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
A Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
A Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
A Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
T The genome sequence of *Schizosaccharomyces pombe*.  
L Nature 415:871-880(2002).  
C -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
C -1- SIMILARITY: TO YEAST SMP3.  
C -1- SIMILARITY: SOME, TO YEAST YGL142C.  
C This SWISS-PROT entry is copyright. It is produced through a collaboration  
C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
C the European Bioinformatics Institute. There are no restrictions on its  
C use by non-profit institutions as long as its content is in no way  
C modified and this statement is not removed. Usage by and for commercial  
C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
C  
C EMBL: Z56276; CAA91213.1; -  
C F.R.; T38857; S62489.  
C GeneDB: Srombe; SPAC4G8.12c; -  
C InterPro: IPR005599; PMP.  
C Pfam: PF03901; PMP; 1.  
C R Hypothetical protein; Transmembrane.  
C T TRANSMEM 8 28 POTENTIAL.  
C T TRANSMEM 61 81 POTENTIAL.  
C T TRANSMEM 91 111 POTENTIAL.  
C T TRANSMEM 144 164 POTENTIAL.  
C T TRANSMEM 175 195 POTENTIAL.  
C T TRANSMEM 216 236 POTENTIAL.  
C T TRANSMEM 274 294 POTENTIAL.  
C T TRANSMEM 297 317 POTENTIAL.  
C T TRANSMEM 338 358 POTENTIAL.  
C T TRANSMEM 496 516 POTENTIAL.  
C T SEQUENCE 533 AA; 62200 MW; F14519C95884687 CRC64;  
Query Match 3.8%; Score 123.5; DB 1; Length 533;  
Best Local Similarity 20.1%; Pred. No. 0.16;  
Matches 112; Conservative 84; Mismatches 180; Indels 181; Gaps 32;  
Y 86 NYWEPHYLYGSGFO-----TWYSPAYAIRSVAYLLHAWPAAPHARILQ 132  
b 26 SYIHPDEHL---OSQIPANKLFGKVKVLPWNEFTYKPIRS-----VWP 66  
Y 133 TNKILVYFRLCLLAFVSCICELFYKAVCKKFG-LHVSRLMLAFVLVS----- 180  
b 67 LNVMLLPFLLC-----RCIC-----KNSCSPYILLFRLVXCLISLLDLSIWNIVPL 116  
Y 181 ----TGNCFCSSAFPLPSF-----CMYTTLIAMTGMVMDKTSIAVLG-----V 219  
b 117 NAKWSALLYSSFMATITQTHFTNSITETIFFTILFLSKNSVPLNKISVLYTFLL 176  
Y 220 AAGAILGW-----PFSALGLP-IAPDLLYKMRWKS-----FFHNSLMALILF-LVPVVV 268  
b 177 AIVSVLGFTRITFLAFVIAPYIYFVSRCKKNVANNPKDIFLHLCIFVSVFATVLACIL 236  
Y 269 IDSYYIKLVIAINIVLVNV-----FTFGPDLYGTPEWYFLINGFLNFNVAFALLLV 324

237 EDYFYGVFVITWNNLKXNSQIENSQHG-----LHSRLTHFTY-----MPLLC 282  
QY 325 LPLTSLMEYLQRFVQNLGHPYWLTLAPMYIWIIFFIQPHKBERFLFPVPLICLOCA 384  
Db 283 GPLI-----FVPLKMDVRK--PATWMLLPVILSLF-----PHQEPFLLPAAISIF---I 328  
QY 385 VALSALOKCY--HFVQRYLEHYVTSNWLALGTVFLGLLSFS----- 427  
Db 329 VNSCLVRSTWIKFLFVY-----AVLAVFGIMHQNGVIAVLEVNKLIEQ 376  
QY 428 RSAV-----LFRGYHGPLDLYPEFYRIATDPT-IHTVPEGRPVNVCKGKWRPSSFL 480  
377 RNVVTMENCLNYPENPETIY--FKKIYSAPTWMALARPKFSQINT---SHLYNYSKQM 431  
QY 481 LPDNLQOLF-----IPSFRRGOLPKPFAAGPLATRIY-PTDMNDQNLSESRXYIDISK 532  
Db 432 ISKFWEYETVEEVKVVNLPB-----NTEFOASTLLVCPVAM-----LQTSYLOQLTM 480  
QY 533 CHYL---VDLDTMEETP 546  
Db 481 LHYIYFHVLDLDDTDELP 497  
RESULT 8  
CYOB\_BUCAI  
ID CYOB\_BUCAI STANDARD; PRT; 662 AA.  
AC P57543;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Ubiquinol oxidase polypeptide 1 (EC 1.10.3.-) (Cytochrome O subunit 1)  
DE (Oxidase 30(3) subunit 1) (Cytochrome O ubiquinol oxidase subunit 1).  
GN CYOB OR BU471.  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
symbiotic bacterium).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Buchnera.  
OX NCBI\_TaxID=118099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tokyo 1998;  
RX MEDLINE=20445173; PubMed=10993077;  
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
RT "Genome sequence of the endocellular bacterial symbiont of aphids  
Buchnera sp. APS";  
RL Nature 407:81-86(2000).  
CC -1- FUNCTION: CYTOCHROME O TERMINAL OXIDASE COMPLEX IS THE COMPONENT  
CC OF THE AEROBIC RESPIRATORY CHAIN THAT PREDOMINATES WHEN CELLS ARE  
CC GROWN AT HIGH AERATION. THIS UBIQUINOL OXIDASE SHOWS PROTON PUMP  
CC ACTIVITY ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER  
CC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: Ubiquinol-8 + O(2) = Ubiquinone-8 + H(2)O.  
CC -1- COFACTOR: CONTAINS TWO PROTOHEME IX (HEME B55 AND B562) AND  
CC COPPER B (BY SIMILARITY).  
CC -1- PATHWAY: UBIQUINOL OXIDASE CATALYZES THE TERMINAL STEP IN THE  
CC ELECTRON TRANSPORT CHAIN.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF001119; BAB13168.1; -  
DR HSPSP; P18401; 1FFT.  
DR InterPro: IPR000883; COX1.  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS; PR01165; CYCOXIDASE1.





QY 385 VALSALQKCYHFVFORYLEHTVTS-NWLALGTAVFLFGLLSFSRVALPRGYHGLDLY 443  
Db 622 DGDUTLQ-----RUKNATPSRDW-----GPALE 646  
QY 444 PFYRIADPTHTVPEGRPVNVCVGEKWTYPPSFLPDNWLQLOPSPFRQOL 498  
Db 647 HRTGRIA--PTIAPSPED-----CFEVSQSLHDPKAQIPIVGSNGSRL 687

RESULT 11  
NUON BUCAI  
ID NUON BUCAI STANDARD; PRT; 469 AA.  
AC P57264;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE NADH-quinone oxidoreductase chain N (EC 1.6.99.5) (NADH dehydrogenase  
I, chain N) (NDH-1, chain N).  
GN NUON OR BUI166  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
symbiotic bacterium).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Buchnera.  
OX NCBI\_TaxID=118099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tokyo 1998;  
RX MEDLINE=20445173; PubMed=10993077;  
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
RT "Genome sequence of the endocellular bacterial symbiont of aphids  
Buchnera sp. APS.";  
RL Nature 407:81-86(2000).  
CC -!- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-  
sulfur (Fe-S) centers, to quinones in the respiratory chain.  
CC Couples the redox reaction to proton translocation (for every two  
electrons transferred, four hydrogen ions are translocated across  
the cytoplasmic membrane), and thus conserves the redox energy in  
a proton gradient (By similarity).  
CC -!- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.  
CC -!- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOA, H, J,  
K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX (BY  
SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AP001118; BAB12884.1; -  
DR InterPro: IPR001750; Oxidored\_q1.  
DR Pfam: PF00361; Oxidored\_q1; 1.  
KW Oxidoreductase; NAD; Quinone; Transmembrane; Complete proteome.  
FT TRANSMEM 19 39 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 87 107 POTENTIAL.  
FT TRANSMEM 108 128 POTENTIAL.  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 184 204 POTENTIAL.  
FT TRANSMEM 252 272 POTENTIAL.  
FT TRANSMEM 281 301 POTENTIAL.  
FT TRANSMEM 312 332 POTENTIAL.  
FT TRANSMEM 359 379 POTENTIAL.  
FT TRANSMEM 381 401 POTENTIAL.  
FT TRANSMEM 442 462 POTENTIAL.  
SQ SEQUENCE 469 AA; 52578 MW; 693705E48E48C62B CRC64;

Query Match 3.5%; Score 113.5; DB 1; Length 469;  
Best Local Similarity 19.7%; Pred. No. 0.73;

Matches 81; Conservative 76; Mismatches 156; Indels 99; Gaps 18;  
QY 126 FHARILQTNKILFYRP-----LRCLAFVSCICELFYKAVCKKGLHVSRRMFLVLVS 180  
Db 37 FLIIIIINVINIIHFNSNAILYIGMILISSICTCTCFSPWLLKYFPNKEEYLLVIIST 96  
QY 181 TGMFCSSAFLPSPFCMYTTLIAMTW-----YMDKTS---IAYLGVAAAGAIL-- 225  
Db 97 LGAISLTISHMASFFINIELISLPMFGLIAYSRYQKYSLESSLKYIILSGVSSPFLFG 156  
QY 226 -GWPFSAALGLPIAFDILLVMKHKWKFHFWMSLMALILFLVPPVVIDSYYVKLVIAPLNI 284  
Db 157 IAWYISISGGL----DFLSI-HKSPFASEKEILVVLFGISMILLSLFF-----KLSI 204  
QY 285 VLYNVFTPHGPDLYGTEPW-----YFYLINGFLNF-----NVAF 318  
Db 205 VPPHLWT---PDYQGSPTSVLSFFSTAGKISVSVLLNLFSLFSNSDNKIVFILSLII 261  
QY 319 ALALLVPLTSLMEYLLQREHVQNLGHPYWLTLAPVWIPIFFIOPHKEERFLPPVPL 378  
Db 262 ILSILVGNLMAFKDKRF-----LGF-----TSIQIGYLLIVLVSHKRYFSLEA-SA 312  
QY 379 ICLOGAVALSALOKCYHFVFORYLEHTVTSNWLALGTAVFLFGLLSFSRVALPRGYH 438  
Db 313 IYLCG-----YLF-----SNIACLGIVNLISTSHINNASSINSYRG 349  
QY 439 PLDLYPEFYRIADPTHTVPEGRPVNVCVGEKWTYPPSFLPDNWLQLOF 489  
Db 350 LFWSHPLLSSVLTLLVLISSA--GIPMTLGFGR--FYILSVIEHLWLIGF 397

RESULT 12  
NUSM TRYBB  
ID NUSM TRYBB STANDARD; PRT; 590 AA.  
AC P04540;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).  
GN ND5.  
OS Trypanosoma brucei brucei.  
OG Mitochondrion.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85037915; PubMed=6093040;  
RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,  
van Boom J.H., Benne R.;  
RT "The sequence of the gene for cytochrome c oxidase subunit I, a  
frameshift containing gene for cytochrome c oxidase subunit II and  
seven unassigned reading frames in Trypanosoma brucei mitochondrial  
maxi-circle DNA.";  
RT Nucleic Acids Res. 12:7327-7344(1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=164;  
RX MEDLINE=87201680; PubMed=2437452;  
RA Jasmier D.P., Feagin J.E., Payne M., Stuart K.;  
RT "Variation of G-rich mitochondrial transcripts among stocks of  
Trypanosoma brucei.";  
RL Mol. Biochem. Parasitol. 22:259-272(1987).  
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X01094; CAB57807.1; -

```

R EMBL; M14820; AAB59225.1; -.
R PIR; A04519; OQUT5.
R InterPro; IPR003916; NADHox redox.
R InterPro; IPR001750; Oxidized q1.
R Pfam; PF00361; oxidized q1; 1.
R PRINTS; PR01434; NADHGNASE5.
W Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
Q SEQUENCE 590 AA; 71494 MW; CAE231AD9DF5282 CRC64;

Query Match 3.4%; Score 112; DB 1; Length 590;
Best Local Similarity 19.5%; Pred. No. 1.2; Indels 186; Gaps 24;
Matches 94; Conservative 65; Mismatches 138;

Y 114 YAYLLHAWPAFAHRIIQTNIKLVYFRLCLLAPV-----SCICEL----- 155
b 93 FCFIVFYAFYMYFDMLLGRFLIIWFVWCVNLFILSYDFLTAYCQWELLGLFSFLLIS 152
Y 156 -FYKAVCKKFGHVRWMLAPLVLTSG---MCSSSAPLPSSFCMYT---TLIAMTG 206
b 153 YWYRPFALKFGK-----AFFIKIGDVLIIAFSIIIFLSNGFCMTTYFLNFFCMY 206
Y 207 WYMDKTSIAVLGVAAGAILGNPFSAAAGL-----PIAFLLVMKHR----- 247
b 207 YXTEFSICLLVGA-----FTKSTQFGLHILWLPDAMEGPIPVSAII--HAATIVVCGII 258
Y 248 -----WKSPEH-----WSMALILFLVPV-----V 267
b 259 LLSFVWCFDFWFSFYNLIGHSTLILMLTLCVFNFDVRYVAFSTICQISFSMECL 318
Y 268 VIDSY-----YGLKVIAPLNLV-----YNYVTPHGPDLVYGP--WYFYL----- 307
b 319 CIDIVIGSLFFCYHMYKATLFIVLGIWHIF-----FGLQDLRCVFFMYFGCVLAR 371
Y 308 -----INGFLNVAFAALILVPLTSLMEVLLQRFHVNQHLGHPYWLTLA 352
b 372 LLILFAILNCSINFLCGLYCKGMILLAL-LMLLSYNIIEFLISI-----IFIFT 422
Y 353 PMYINPIIPIOPHKEERLFPVYPLI-CL-----CGAVALSALQKV----- 394
b 423 MIYNYELLFFLM-----FVFKCFCLVDCFLFLPDYECVCLVCLISLYMCILSIFFIIDF 476
Y 395 -----HFVQRYLEHYTVTSNWLALGTVPGLLSFSRSVALRGVHGPDLDLYPEF 446
b 477 VCIFFPSSCYVFWFFNPFNFDF--IAIFVFLILSVGLFYGLCF-FYFFNIDCIMLF 533
Y 447 YRI 449
b 534 WRI 536

RESULT 13
AGG ECOLI
D -YAGG ECOLI STANDARD; PRT; 460 AA.
P75683; P71292;
T 01-NOV-1997 (Rel. 35, Created)
T 01-NOV-1997 (Rel. 35, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Hypothetical symporter yagG.
N YAGG OR B0270.
S Escherichia coli.
S Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
S Enterobacteriaceae; Escherichia.
X NCBI_TaxID=562;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=K12 / MG1655;
X MEDLINE=97426617; PubMed=9278503;
A Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
A Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
A Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
A Mau B., Shao Y.;
T "The complete genome sequence of Escherichia coli K-12."
Science 277:1453-1474 (1997).

```

```

[2]
RN SEQUENCE FROM N.A.
RP Schramm S., Duncan M., Allen E., Ataujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC (SGF). STRONG, TO E.COLI YICU.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000135; AAC73373.1; -.
DR EMBL; U70214; AAB08691.1; -.
DR PIR; F64752; F64752.
DR EcoGene; EG13346; yagG.
DR InterPro; IPR001927; Na/Gal symport.
DR TIGRFAMs; TIGR00792; gph; 1.
DR PROSITE; PS00872; NA GALACTOSIDE SYMP; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Symport; Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 309 329 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT CONFLICT 370 370 K -> N (IN REF. 2).
SQ SEQUENCE 460 AA; 50631 MW; AB600589E663C6B7 CRC64;

Query Match 3.4%; Score 111; DB 1; Length 460;
Best Local Similarity 22.7%; Pred. No. 1.1; Indels 82; Gaps 16;
Matches 67; Conservative 46; Mismatches 100;

QY 195 FCMYTTLIAMTGWYMDKTSIAVLGVAAGAILGNPFSAAAGLPIAFD-----LLVMKHR--- 247
Db 21 FVWQATMFLAYFYTD-----VFGLSAG-IMGTLFLVSRVLDVATDPLMGLLVDRTRRH 74
QY 248 --WKSPEH-----SLMALILFLVPVVVIDSYYYKLVIAPIALVLY----- 287
Db 75 GQFRPFLWGAIPFGIVCVLTFYTP-----DFSAGKIIYACVYVILTLVYTFVNPYCA 130
QY 288 --NVFTPHGPDLYGTEPWYFYLYNGFLNFNVAFAAL--LVPLTSLM-----EYLLQRP 339
Db 131 MFGVITADPKERHALQSWRFELA-----AAGSLAISGIALPLVSIIGKGE-----Q 177
QY 340 VQNLGHPYWLILAPYIWFIIFFIOPHKEERLFPVY-----LILCG-----AVA 386
Db 178 VGYFGAMCVLGLSGVILLYVCFET---TKERYTEVQPGSSVAKDLKLLGNSQWRIMCA 234
QY 387 LSAQKQCVHFV-----FQYRLEHYTVTSNWLALGTV-FLFGLLSFSRSVALF 433
Db 235 FQWATCSNVVRGGATLVFYKVVMDHPELATQFLLYGLSLATMFGSLCSSLGLGPF 289

RESULT 14
Y048_UREPA
ID Y048_UREPA STANDARD; PRT; 496 AA.
AC Q9PR37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```
16-OCT-2001 (Rel. 40, Last annotation update)
Hypothetical protein UU048.
Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
[1]
SEQUENCE FROM N.A.
STRAIN=Serovar 3; PubMed=11048724;
MEDLINE=30500219; Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
Cassell G.H.;
"Ureaplasma urealyticum."
Nature 407:757-762(2000).
The complete sequence of the mucosal pathogen Ureaplasma
urealyticum."
[1]
SUBCELLULAR LOCATION: Integral membrane protein (Potential).
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; AE002104; AAF30453.1;
Hypothetical protein; Transmembrane; Complete proteome.
TRANSMEM 33 53 POTENTIAL.
TRANSMEM 89 109 POTENTIAL.
TRANSMEM 127 147 POTENTIAL.
TRANSMEM 154 174 POTENTIAL.
TRANSMEM 193 213 POTENTIAL.
TRANSMEM 247 267 POTENTIAL.
TRANSMEM 285 305 POTENTIAL.
TRANSMEM 320 340 POTENTIAL.
TRANSMEM 355 375 POTENTIAL.
TRANSMEM 382 402 POTENTIAL.
TRANSMEM 411 431 POTENTIAL.
TRANSMEM 455 475 POTENTIAL.
SEQUENCE 496 AA; 59349 MW; 4F2330427DC3525D CRC64;
Query Match 3.4%; Score 110.5; DB 1; Length 496;
Best Local Similarity 19.8%; Pred. No. 1.3;
Matches 78; Conservative 63; Mismatches 131; Indels 121; Gaps 17;
95 IYGGFQWEYSPAYATRSYAVYLLHAWPAAPFARHILQTNKILVFLRCLAF----- 148
57 LFLGFGDENKSMISKSLNAPFDLHS-----PKYLQNLFLVFRFP--ILSFLFTYL 108
149 VSCICELYFYKAVCKFGLHVSRLAFVLSTGMPFCSSAFPLPSFCMYTTLIAMTGWY 208
109 IKNFTNLYWHRATIKKY---LPWFVLYLVIA-----ISFLFFTFFSV---- 149
209 MDKTSIAVLGVAGAILGWPSPSALGFIAPFDLLVMKRWKSFHWSLWALILFLVPV-V 267
150 -----WP-----KEVFNVLVLLVIFLNLISY 171
268 VIDSYY-----YGLKVIAPL---NIVLYNVFTPHGPDLYGTFFPYFYLINGFL 312
172 EIFNYFISKTNPLLYDNYKNLIAMVFOALLLFLVITPLVWINTGKSNFLVDNRFY 231
313 -----NFVA-----FALALVPLTSLMEVLLQRFVQV-LGHPYV---LT 350
232 TRIVDITVQSGKRNFIILIAFFFLITFIIVLANTNPFALVINKRDYRNKKNLFIILL 291
351 LAPMYIWFIFPIQPHKEERF-----LFPVYPLICLCAVAL-----SALQKCYHFVFQ 399
292 PSAIFIMLLRVFAYKHENENLPVGNHLLWVYLQSFPAIILLYWVTLKRLSVKSS 351
400 RYRLEHTVTSNWLALGTFLFGLLSFRSVAL 432
352 LNTLLNVLVVTQITLSL-SLFLVTLFNSKSVSL 383
```

RESULT 15  
S6A9\_MOUSE STANDARD; PRT; 633 AA.  
ID S6A9\_MOUSE  
AC P28571; (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last annotation update)  
DT 01-DEC-1992 (Rel. 24, Last annotation update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Sodium- and chloride-dependent glycine transporter 1 (GlyT1) (GlyT-1).  
GN SLC6A9 OR GLYT1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SOURCE FROM N.A. (ISOFORMS GLYT-1A AND GLYT-1B).  
RC TISSUE=Brain;  
RX MEDLINE=92316254; PubMed=1618338;  
RA Liu Q.-R., Nelson H., Mandiyan S., Lopez-Corcuera B., Nelson N.;  
RT "Cloning and expression of a glycine transporter from mouse brain."  
RL FEBS Lett. 305:110-114(1992).  
CC -!- FUNCTION: Terminates the action of glycine by its high affinity  
CC sodium-dependent reuptake into presynaptic terminals. May play a  
CC role in regulation of glycine levels in NMDA receptor-mediated  
CC neurotransmission.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=GlyT-1A;  
CC IsoId=P28571-1; Sequence=Displayed;  
CC Name=GlyT-1B;  
CC IsoId=P28571-2; Sequence=VSP\_006272;  
CC -!- SIMILARITY: Belongs to the sodium:neurotransmitter symporter (SNF)  
CC family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; X67056; CAA47440.1; -.  
PIR; S23151; S23151.  
MD; MGI:95760; SLC6A9.  
InterPro; IPR000175; Na/ntran\_symport.  
Pfam; PF00209; SNF; 1.  
PRINTS; PR00176; NANEUSMPORT.  
ProDom; PD00448; Na/ntran\_symport; 2.  
PROSITE; PS00610; NA\_NEUOTRAN\_SYMP\_1; 1.  
PROSITE; PS00754; NA\_NEUOTRAN\_SYMP\_2; 1.  
PROSITE; PS00754; NA\_NEUOTRAN\_SYMP\_3; 1.  
PROSITE; PS00267; NA\_NEUOTRAN\_SYMP\_3; 1.  
Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;  
Symport; Amino-acid transport; Alternative splicing.  
DOMAIN 1 35 CYTOPLASMIC (POTENTIAL).  
TRANSMEM 36 56 1 (POTENTIAL).  
TRANSMEM 64 83 2 (POTENTIAL).  
TRANSMEM 107 127 3 (POTENTIAL).  
DOMAIN 128 214 EXTRACELLULAR (POTENTIAL).  
TRANSMEM 215 233 4 (POTENTIAL).  
TRANSMEM 242 259 5 (POTENTIAL).  
TRANSMEM 295 312 6 (POTENTIAL).  
TRANSMEM 324 345 7 (POTENTIAL).  
TRANSMEM 378 397 8 (POTENTIAL).  
TRANSMEM 426 444 9 (POTENTIAL).  
TRANSMEM 460 480 10 (POTENTIAL).  
TRANSMEM 501 520 11 (POTENTIAL).  
TRANSMEM 539 557 12 (POTENTIAL).  
DOMAIN 558 633 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 164 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 N-LINKED (GLCNAC. . .) (POTENTIAL).

Search completed: January 13, 2004, 07:02:08  
Job time : 24 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 13, 2004, 06:32:04 ; Search time 59 Seconds

(without alignments)

2672.375 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGASSGDT.....VNYTLKPRKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvrius:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3274	100.0	611	4 Q9H068	Q9h068 homo sapien
2	3268	99.8	611	4 Q96P89	Q96pb9 homo sapien
3	3260.5	99.6	618	4 Q96G57	Q96g87 homo sapien
4	3259.5	99.6	618	4 Q9H6U8	Q9h6u8 homo sapien
5	3043	92.9	611	11 Q8VDI9	Q8vdi9 mus musculus
6	3039	92.8	611	11 Q8C7G0	Q8c7g0 mus musculus
7	2739	83.7	570	11 Q8C378	Q8c378 mus musculus
8	1358	41.5	623	5 Q9VBV8	Q9vbv8 drosophila
9	997	30.5	198	6 Q9BGV0	Q9bgv0 macaca fasc
10	849.5	25.9	596	10 Q8LRH5	Q8lrh5 oryza sativ
11	812	24.8	570	10 Q9FZ49	Q9fz49 arabidopsis
12	376	11.5	78	11 Q8BT44	Q8bt44 mus musculus
13	296.5	9.1	548	10 Q9A1L5	Q9a1l5 arabidopsis
14	278.5	8.5	498	10 Q9LEQ5	Q9leg5 arabidopsis
15	264.5	7.2	496	5 Q967F2	Q967f2 caenorhabdi
16	235	7.2	561	5 Q9VZM5	Q9vzm5 drosophila

17	234	7.1	506	10 Q9AWW0	Q9aww0 oryza sativ
18	219.5	6.7	558	5 Q9NKKZ7	Q9nkz7 trypanosoma
19	209.5	6.4	554	4 Q92521	Q92521 homo sapien
20	208.5	6.4	554	4 Q8WVN7	Q8wvn7 homo sapien
21	207	6.3	506	3 Q9USN0	Q9usn0 schizosacch
22	194.5	5.9	542	11 Q9JQ0	Q9jq0 mus musculus
23	193	5.9	492	5 Q23361	Q23361 caenorhabdi
24	183	5.6	509	2 Q9RNP1	Q9rnp1 zymomonas m
25	182.5	5.6	488	4 Q96AA4	Q96aa4 homo sapien
26	181.5	5.5	488	4 Q9BV10	Q9bv10 homo sapien
27	181.5	5.5	683	4 Q8NG10	Q8ng10 homo sapien
28	180.5	5.5	678	5 Q9VH78	Q9vh78 drosophila
29	168	5.1	696	5 Q8MT80	Q8mt80 drosophila
30	168	5.1	830	5 Q9W176	Q9w176 drosophila
31	161	4.9	483	11 Q8VDB2	Q8vdb2 mus musculu
32	141	4.3	547	3 Q74753	Q74753 schizosacch
33	135.5	4.1	603	2 Q8KJD8	Q8kjd8 rhizobium l
34	132.5	4.0	491	2 Q9EV68	Q9ev68 rhizobium m
35	131.5	4.0	487	16 Q9JS51	Q9js51 chlamydia p
36	131.5	4.0	487	16 Q929E6	Q929e6 chlamydia p
37	131.5	4.0	491	16 Q92YNE	Q92yn6 rhizobium m
38	130.5	4.0	583	16 Q9FEU9	Q9feu9 xylella fas
39	130	4.0	433	16 P1970	P1970 mycobacteri
40	129	3.9	477	16 Q8YB02	Q8yb02 brucella me
41	127.5	3.9	421	2 Q9L9P3	Q9l9p3 lactococcus
42	127	3.9	465	17 Q50564	Q50564 methanobact
43	127	3.9	537	16 Q98LJ8	Q98lj8 rhizobium l
44	127	3.9	718	17 Q97ZT2	Q97zt2 sulfolobus
45	126.5	3.9	1962	6 Q46669	Q46669 canis famil

ALIGNMENTS

RESULT 1

Q9H068	PRELIMINARY;	PRT;	611 AA.
ID Q9H068			
AC Q9H068;			
DT 01-MAR-2001 (TREMBLrel. 16, Created)			
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE Hypothetical protein.			
GN DKFZP586M2420.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Uterus;			
RX MEDLINE=21154917; PubMed=11230166;			
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,			
RA Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,			
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,			
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,			
RA Wamburt R., Korn B., Klein M., Poustka A.;			
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and			
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."			
RL Genome Res. 11:422-435(2001).			
DR EMBL; AL136927; CAB66861.1; -			
KW Hypothetical protein.			
SQ SEQUENCE 611 AA; 69863 MW; 51EC72DDB866713 CRC64;			

Query Match	100.0%;	Score 3274;	DB 4;	Length 611;
Best Local Similarity	100.0%;	Pred. No. 3.7e-257;		
Matches 611;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 MASRGARQLKSGASSGDTAPADKRLRLSREAGGAHRTLSGNKAGQVWPEGST 60			
Db	1 MASRGARQLKSGASSGDTAPADKRLRLSREAGGAHRTLSGNKAGQVWPEGST 60			
QY	61 AFKLLSARLCALLSNISDCDETNWPEPTHLYLXGEGFTWEXSPAYRISYVLLIH 120			

```

61 AFKLLSARLCAALLSNISDCDETFNWEPTHLYIYGEQFTWEYSPAVAIRSVAYLLH 120
121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
181 TGMFCSSAFILPSSFCMYTTLIANTGMYMDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
181 TGMFCSSAFILPSSFCMYTTLIANTGMYMDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
241 LLVMKRWKSFHWSLMALILFLVPPVVVWIDSYGKLVIAPLNVLVYNVFTPHGPDLYGT 300
241 LLVMKRWKSFHWSLMALILFLVPPVVVWIDSYGKLVIAPLNVLVYNVFTPHGPDLYGT 300
301 EPWYFYLLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLPAMYIWFII 360
301 EPWYFYLLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLPAMYIWFII 360
361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCYHFVQRYRLEHYTIVTSNMLALGTVEL 420
361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCYHFVQRYRLEHYTIVTSNMLALGTVEL 420
421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
481 LPDNWQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
481 LPDNWQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
541 TMRETPREPKYSSNKEEWISLAYRPFDDASRSKLLRAFYVPFLSDQTYVYVNTILKPR 600
541 TMRETPREPKYSSNKEEWISLAYRPFDDASRSKLLRAFYVPFLSDQTYVYVNTILKPR 600
601 KAKQIRKXSGG 611
601 KAKQIRKXSGG 611

RESULT 2
96PB9 PRELIMINARY; PRT; 611 AA.
AC Q96PB9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DIBD1.
E DIBD1.
S Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
KW "DIBD1, a novel gene homologous to Saccharomyces cerevisiae ALG9, is
translocated in a family with bipolar affective disorder by a
translocation breakpoint at 11q23."
L Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; AF39532; AAL25798.1; -.
Q SEQUENCE 611 AA; 69877 MW; 2736B6750B5ED302 CRC64;

Query Match 99.8%; Score 3268; DB 4; Length 611;
Best Local Similarity 99.8%; Pred. No. 1.1e-256;
Matches 610; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MASRGARQRLKSGSGSGTAPAAKLLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
1 MASRGARQRLKSGSGSGTAPAAKLLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
61 AFKLLSARLCAALLSNISDCDETFNWEPTHLYIYGEQFTWEYSPAVAIRSVAYLLH 120

```

```

Db 61 AFKLLSARLCAALLSNISDCDETFNWEPTHLYIYGEQFTWEYSPAVAIRSVAYLLH 120
Qy 121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
Db 121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
Qy 181 TGMFCSSAFILPSSFCMYTTLIANTGMYMDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
Db 181 TGMFCSSAFILPSSFCMYTTLIANTGMYMDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
Qy 241 LLVMKRWKSFHWSLMALILFLVPPVVVWIDSYGKLVIAPLNVLVYNVFTPHGPDLYGT 300
Db 241 LLVMKRWKSFHWSLMALILFLVPPVVVWIDSYGKLVIAPLNVLVYNVFTPHGPDLYGT 300
Qy 301 EPWYFYLLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLPAMYIWFII 360
Db 301 EPWYFYLLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLPAMYIWFII 360
Qy 361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCYHFVQRYRLEHYTIVTSNMLALGTVEL 420
Db 361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCYHFVQRYRLEHYTIVTSNMLALGTVEL 420
Qy 421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
Db 421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
Qy 481 LPDNWQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
Db 481 LPDNWQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
Qy 541 TMRETPREPKYSSNKEEWISLAYRPFDDASRSKLLRAFYVPFLSDQTYVYVNTILKPR 600
Db 541 TMRETPREPKYSSNKEEWISLAYRPFDDASRSKLLRAFYVPFLSDQTYVYVNTILKPR 600
Qy 601 KAKQIRKXSGG 611
Db 601 KAKQIRKXSGG 611

RESULT 3
96GS7 PRELIMINARY; PRT; 618 AA.
AC Q96GS7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
KW EMBL; BC009255; AAH09255.1; -.
KW Hypothetical protein.
SQ SEQUENCE 618 AA; 70786 MW; 05790C3513BF8A4D CRC64;

Query Match 99.6%; Score 3260.5; DB 4; Length 618;
Best Local Similarity 98.9%; Pred. No. 4.7e-256;
Matches 611; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

1 MASRGARQRLKSGSGSGTAPAAKLLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
1 MASRGARQRLKSGSGSGTAPAAKLLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
61 AFKLLSARLCAALLSNISDCDETFNWEPTHLYIYGEQFTWEYSPAVAIRSVAYLLH 120
61 AFKLLSARLCAALLSNISDCDETFNWEPTHLYIYGEQFTWEYSPAVAIRSVAYLLH 120
121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180

```

121 AWPAAFHARILQTNKILVFFLRCLLAFVSCICELFYKAVCKFGLHVSRRMLAFVLVS 180  
181 TGMFCSSSAFLPSSFCMYTTLIAMTGWMDKTSIAVLGVAAGAILGWFFSAALGLPIAFD 240  
181 TGMFCSSSAFLPSSFCMYTTLIAMTGWMDKTSIAVLGVAAGAILGWFFSAALGLPIAFD 240  
241 LLMVGRHWSKSPHWSLMALILFLVPVVVWIDSYGKLVIAPIANIVLVNVPHPGPDLYGT 300  
241 LLMVGRHWSKSPHWSLMALILFLVPVVVWIDSYGKLVIAPIANIVLVNVPHPGPDLYGT 300  
301 EPWYFYLINGFLNFVAFALALVPLTSLMEYLLQRFHVONLGHYPWLTLPAPWYIWFII 360  
301 EPWYFYLINGFLNFVAFALALVPLTSLMEYLLQRFHVONLGHYPWLTLPAPWYIWFII 360  
361 FFIQPHKEERFLFPVYPLICLGAVALSAL-----QKCVHFVQRYRLEHYTTSNWL 413  
361 FFIQPHKEERFLFPVYPLICLGAVALSALQHSFLYFKCVHFVQRYRLEHYTTSNWL 420  
414 ALGTVFLFGLLSFSRSVALFRGYHGLDLYPEFYRIATDPTIHTVPEGRPVNVCVKEMY 473  
421 ALGTVFLFGLLSFSRSVALFRGYHGLDLYPEFYRIATDPTIHTVPEGRPVNVCVKEMY 480  
474 RFPSSFLPDNWLQFIPEFRGQLPKPAEGPLATRIVPTDMNDQNLLEPSRYIDISK 533  
481 RFPSSFLPDNWLQFIPEFRGQLPKPAEGPLATRIVPTDMNDQNLLEPSRYIDISK 540  
534 HYLVDLDTWRETREPKYSSNKEEWSLAYRPFLLDASRSSKLLRAFYVPLSDQYTVVYN 593  
541 HYLVDLDTWRETREPKYSSNKEEWSLAYRPFLLDASRSSKLLRAFYVPLSDQYTVVYN 600  
594 YTIKPRKAKQIRKXSGG 611  
601 YTIKPRKAKQIRKXSGG 618

RESULT 4  
Q9H6U8 PRELIMINARY; PRT; 618 AA.  
AC Q9H6U8;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Similar to disrupted in bipolar disorder 1 (DIBD1 homolog).  
GN 8230402H15RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,  
RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;  
RA "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
JR EMBL; AK025498; BAB15154.1; -;  
GW Hypothetical protein.  
SQ SEQUENCE 618 AA; 70800 MW; A2EAPAB7FAS5F6BD CRC64;

Query Match 99.6%; Score 3259.5; DB 4; Length 618;  
Best Local Similarity 98.7%; Pred. No. 5.6e-256;  
Matches 610; Conservative 1; Mismatches 0; Indels 7; Gaps 1;  
1 MASRGARQLKSGASSGDTAPADKRLRELLGSREAGGAHRTLSGNKAGQWAPGST 60  
1 MASRGARQLKSGASSGDTAPADKRLRELLGSREAGGAHRTLSGNKAGQWAPGST 60  
61 AFKLLSARLCAALLSNISDCDETFNYWETHYLYIGEFTWYGPAYAIRSYAYLLH 120  
61 AFKLLSARLCAALLSNISDCDETFNYWETHYLYIGEFTWYGPAYAIRSYAYLLH 120  
121 AWPAAFHARILQTNKILVFFLRCLLAFVSCICELFYKAVCKFGLHVSRRMLAFVLVS 180

Db 121 AWPAAFHARILQTNKILVFFLRCLLAFVSCICELFYKAVCKFGLHVSRRMLAFVLVS 180  
QY 181 TGMFCSSSAFLPSSFCMYTTLIAMTGWMDKTSIAVLGVAAGAILGWFFSAALGLPIAFD 240  
Db 181 TGMFCSSSAFLPSSFCMYTTLIAMTGWMDKTSIAVLGVAAGAILGWFFSAALGLPIAFD 240  
QY 241 LLMVGRHWSKSPHWSLMALILFLVPVVVWIDSYGKLVIAPIANIVLVNVPHPGPDLYGT 300  
Db 241 LLMVGRHWSKSPHWSLMALILFLVPVVVWIDSYGKLVIAPIANIVLVNVPHPGPDLYGT 300  
QY 301 EPWYFYLINGFLNFVAFALALVPLTSLMEYLLQRFHVONLGHYPWLTLPAPWYIWFII 360  
Db 301 EPWYFYLINGFLNFVAFALALVPLTSLMEYLLQRFHVONLGHYPWLTLPAPWYIWFII 360  
QY 361 FFIQPHKEERFLFPVYPLICLGAVALSAL-----QKCVHFVQRYRLEHYTTSNWL 413  
Db 361 FFIQPHKEERFLFPVYPLICLGAVALSALQHSFLYFKCVHFVQRYRLEHYTTSNWL 420  
QY 414 ALGTVFLFGLLSFSRSVALFRGYHGLDLYPEFYRIATDPTIHTVPEGRPVNVCVKEMY 473  
Db 421 ALGTVFLFGLLSFSRSVALFRGYHGLDLYPEFYRIATDPTIHTVPEGRPVNVCVKEMY 480  
QY 474 RFPSSFLPDNWLQFIPEFRGQLPKPAEGPLATRIVPTDMNDQNLLEPSRYIDISK 533  
Db 481 RFPSSFLPDNWLQFIPEFRGQLPKPAEGPLATRIVPTDMNDQNLLEPSRYIDISK 540  
QY 534 HYLVDLDTWRETREPKYSSNKEEWSLAYRPFLLDASRSSKLLRAFYVPLSDQYTVVYN 593  
Db 541 HYLVDLDTWRETREPKYSSNKEEWSLAYRPFLLDASRSSKLLRAFYVPLSDQYTVVYN 600  
QY 594 YTIKPRKAKQIRKXSGG 611  
Db 601 YTIKPRKAKQIRKXSGG 618

RESULT 5  
Q8VDI9 PRELIMINARY; PRT; 611 AA.  
AC Q8VDI9;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Similar to disrupted in bipolar disorder 1 (DIBD1 homolog).  
GN 8230402H15RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Ovary;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium.  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; BC021791; AAH21791.1; -;  
DR EMBL; AK054293; BAC35720.1; -;  
DR MGD; MGI:1924753; 8230402H15Rik.  
SQ SEQUENCE 611 AA; 69560 MW; 3C96FFC7854F0133 CRC64;

Query Match 92.9%; Score 3043; DB 11; Length 611;  
Best Local Similarity 93.5%; Pred. No. 2e-238;  
Matches 574; Conservative 9; Mismatches 25; Indels 6; Gaps 4;  
1 MASRGARQLKSGASSGDTAPADKRLRELLGSREAGGAHRTLSGNKAGQWAPGST 57  
1 MASRGARQLKSGASSGDTAPADKRLRELLGSREAGGAHRTLSGNKAGQWAPGST 57

58 GSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYVGEQOTWEYSPAYAIRSYAYL 117  
 58 GSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYVGEQOTWEYSPAYAIRSYAYL 117  
 118 LHWAPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHVSRRMLAFL 177  
 118 LHWAPAAFHARILQTNKILVFLRCLLAFVSCVCELYFYKAVCKKFGHVSRRMLAFL 177  
 178 VLSTGMFCSSSAFLPSSFCMYTTLIAMTGMNDKTSIAVLGVAAGAILGWPFSAALGLPI 237  
 178 VLSTGMFCSSSAFLPSSFCMYTTLIAMTGMNDKTPIAVLGVAAGAILGWPFSAALGLPI 237  
 238 AFDLLVMKRWKSFHSLMALILFLVPVVVDSYYGKLVIAPLNVLVYNVFTSHGPD 297  
 238 AFDLLVMKRWKSFHSLMALILFLVPVVVDSYYGKLVIAPLNVLVYNVFTSHGPD 297  
 298 YGTEPWYFLINGFLNFVAFALALLVPLTSLMEYLLQRFHVONLGHYPWLTAPMYIW 357  
 298 YGTEPWYFLINGFLNFVAFALALLVPLTSLMEYLLQRFHVONLGHYPWLTAPMYIW 357  
 358 FIIFIQPHKEERFLFPVPLICGVALSALQKCYHFVQRYLEHYTVTSNWLALGT 417  
 358 FIIFIQPHKEERFLFPVPLICGVALSALQKCYHFVQRYLEHYTVTSNWLALGT 417  
 418 VFLGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCCKEWRPFS 477  
 418 VFLGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCCKEWRPFS 477  
 478 SFLLPDNWLQFIPSEFGQPKPFAEGPLATRVPTDMNDQNLBEPSSRYIDISKCHYL 537  
 478 SFLLPDNWLQFIPSEFGQPKPFAEGPLATRVPTDMNDQNLBEPSSRYIDISKCHYL 537  
 538 DLDTMRETPREPKNKEEWSIAVLPFLDASRSSKLLAFVYVPPFSDQYTVVNYTIL 597  
 538 DLDTMRETPREPKNKEEWSIAVLPFLDASRSSKLLAFVYVPPFSDQYTVVNYTIL 597  
 598 KPRKAKQIRKXSGG 611  
 598 KPRKAKQIRKXSGG 611

RESULT 6  
 BC7G0 PRELIMINARY; PRT; 611 AA.  
 C Q8C7G0;  
 D Q8C7G0;  
 T 01-MAR-2003 (Tremblrel. 23, Created)  
 T 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 T 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 E DIBD1 homolog.  
 S Mus musculus (Mouse).  
 C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 X NCBI\_TaxID=10090;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=C57BL/6J; TISSUE=Liver;  
 X MEDLINE=22354683; PubMed=12466851;  
 A The FANTOM Consortium.  
 A the RIKEN Genome Exploration Research Group Phase I & II Team;  
 T "Analysis of the mouse transcriptome based on functional annotation of  
 T 60,770 full-length cDNAs."  
 L Nature 420:563-573 (2002).  
 R EMBL; AK050335; BAC34195.1; --  
 Q SEQUENCE 611 AA; 69588 MW; 6298C0C6DOBBI742 CRC64;

Query Match 92.8%; Score 3039; DB 11; Length 611;  
 Best Local Similarity 93.3%; Pred. No. 4.3e-238;  
 Matches 573; Conservative 10; Mismatches 25; Indels 6; Gaps 4;  
 Y 1 MASRRARORLK---GSGASSGDTAPAADKRLSRRAGGAEHTELSGNKAGQVWAPE 57  
 b 1 MASRRARORLKGSGGGGGGGGGGAGPAAEKL-EQLGSREA-GAEPRPE-SGNKAGQVWAPE 57

QY 58 GSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYVGEQOTWEYSPAYAIRSYAYL 117  
 DB 58 GSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYVGEQOTWEYSPAYAIRSYAYL 117  
 QY 118 LHWAPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHVSRRMLAFL 177  
 DB 118 LHWAPAAFHARILQTNKILVFLRCLLAFVSCVCELYFYKAVCKKFGHVSRRMLAFL 177  
 QY 178 VLSTGMFCSSSAFLPSSFCMYTTLIAMTGMNDKTSIAVLGVAAGAILGWPFSAALGLPI 237  
 DB 178 VLSTGMFCSSSAFLPSSFCMYTTLIAMTGMNDKTPIAVLGVAAGAILGWPFSAALGLPI 237  
 QY 238 AFDLLVMKRWKSFHSLMALILFLVPVVVDSYYGKLVIAPLNVLVYNVFTSHGPD 297  
 DB 238 AFDLLVMKRWKSFHSLMALILFLVPVVVDSYYGKLVIAPLNVLVYNVFTSHGPD 297  
 QY 298 YGTEPWYFLINGFLNFVAFALALLVPLTSLMEYLLQRFHVONLGHYPWLTAPMYIW 357  
 DB 298 YGTEPWYFLINGFLNFVAFALALLVPLTSLMEYLLQRFHVONLGHYPWLTAPMYIW 357  
 QY 358 FIIFIQPHKEERFLFPVPLICGVALSALQKCYHFVQRYLEHYTVTSNWLALGT 417  
 DB 358 FIIFIQPHKEERFLFPVPLICGVALSALQKCYHFVQRYLEHYTVTSNWLALGT 417  
 QY 418 VFLGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCCKEWRPFS 477  
 DB 418 VFLGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCCKEWRPFS 477  
 QY 478 SFLLPDNWLQFIPSEFGQPKPFAEGPLATRVPTDMNDQNLBEPSSRYIDISKCHYL 537  
 DB 478 SFLLPDNWLQFIPSEFGQPKPFAEGPLATRVPTDMNDQNLBEPSSRYIDISKCHYL 537  
 QY 538 DLDTMRETPREPKNKEEWSIAVLPFLDASRSSKLLAFVYVPPFSDQYTVVNYTIL 597  
 DB 538 DLDTMRETPREPKNKEEWSIAVLPFLDASRSSKLLAFVYVPPFSDQYTVVNYTIL 597  
 QY 598 KPRKAKQIRKXSGG 611  
 DB 598 KPRKAKQIRKXSGG 611

RESULT 7  
 BC378 PRELIMINARY; PRT; 570 AA.  
 ID Q8C378;  
 AC Q8C378;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE DIBD1 homolog.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NCBI\_TaxID=10090;  
 RN [1]  
 P SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium.  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 DR EMBL; AK086674; BAC39717.1; --  
 SQ SEQUENCE 570 AA; 64393 MW; 8FFAE5572461A117 CRC64;

Query Match 83.7%; Score 2739; DB 11; Length 570;  
 Best Local Similarity 93.5%; Pred. No. 8.6e-214;  
 Matches 517; Conservative 5; Mismatches 25; Indels 6; Gaps 4;  
 QY 1 MASRRARORLK---GSGASSGDTAPAADKRLSRRAGGAEHTELSGNKAGQVWAPE 57  
 DB 1 MASRRARORLKGSGGGGGGGGGGAGPAAEKL-EQLGSREA-GAEPRPE-SGNKAGQVWAPE 57

```
QY 58 GSTAFKCLLSARLCAALLSNISDCDETFNYWEPYTHLYIGEGFQWTSYPAIRSVAYL 117
Db 58 GSTAFKCLLSARLCAALLSNISDCDETFNYWEPYTHLYIGEGFQWTSYPAIRSVAYL 117
QY 118 LHAWPAFAHARILQTNKILVFLRCLLAFVSCICELYFVKACKFGLHVRMMLAF 177
Db 118 LHAWPAFAHARILQTNKILVFLRCLLAFVSCICELYFVKACKFGLHVRMMLAF 177
QY 178 VLSGTGFCSSSAFLPSSFCMTYLLIAMTGMWMDKTSIAVLGVAAGAILGWPFAALGLPI 237
Db 178 VLSGTGFCSSSAFLPSSFCMTYLLIAMTGMWMDKTSIAVLGVAAGAILGWPFAALGLPI 237
QY 238 AFDLLVMKRWKSPFHSNLSMALILFLVPVVIDSYTYGKLVIAPLNVLNVFTPHGPD 297
Db 238 AFDLLVMKRWKSPFHSNLSMALILFLVPVVIDSYTYGKLVIAPLNVLNVFTPHGPD 297
QY 298 YCTEPWYFYLLNGFLNFVAFALALLVPLTSLMEYLLQRFVONLGHYPWYLLAPYIW 357
Db 298 YCTEPWYFYLLNGFLNFVAFALALLVPLTSLMEYLLQRFVONLGHYPWYLLAPYIW 357
QY 358 FIIFIQPKERFLFPVYPICLCGAVALSALOKCHYFQYRLHYVTWGNALGT 417
Db 358 FIIFIQPKERFLFPVYPICLCGAVALSALOKCHYFQYRLHYVTWGNALGT 417
QY 418 VFLFGLLSFSRVALFRGVHGLDLYPEFYRIADPTHTVPEGRPVNVCWKWYRFP 477
Db 418 VFLFGLLSFSRVALFRGVHGLDLYPEFYRIADPTHTVPEGRPVNVCWKWYRFP 477
QY 478 SFLLPDNMQLOPISFEGQLPKPPAEGPLATRIPTMDNDQNLLEPSRYIDISKCHYL 537
Db 478 SFLLPDNMQLOPISFEGQLPKPPAEGPLATRIPTMDNDQNLLEPSRYIDISKCHYL 537
QY 538 DLDWTRETPREP 550
Db 538 DLDWTRETPREP 550

RESULT 8
QYVB8 PRELIMINARY; PRT; 623 AA.
ID Q9VB8 STRAIN=BERKELEY;
AC Q9VB8
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 20, Last annotation update)
DE CG11851 protein (LD40966p).
GN CG11851
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottker P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke K., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glisdek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
```

```
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2183-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Pargass V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003751; AAF56419.1; -
DR EMBL; AY069643; AAL39788.1; -
DR FlyBase; FBgn0039293; CG11851.
SQ SEQUENCE 623 AA; 72009 MW; 97FA5120BD432850 CRC64;

Query Match 41.5%; Score 1358; DB 5; Length 623;
Best Local Similarity 42.4%; Pred. No. 1e-101;
Matches 267; Conservative 115; Mismatches 213; Indels 34; Gaps 10;

QY 1 MASRGARQLKGGSSGDTAPAADKLRELLG---SREAGGAHRTELSTGNKAGQWA-- 55
Db 1 MAPFAARARVIANKADNQLSKPPKPPGLNGNNKTKYATPAGKKDKDAKRNQPTSG 60
QY 56 -----PEGSTAPKCLLSARLCAALLSNISDCDETFNYWEPYTHLYIGEGFQWTSY 105
Db 61 QEKGLPNPIMPSTQPTAFKTVSARLCSAIWAYIADCDTFNYWEPYTHLYINGHGLTWEY 120
QY 106 SPAYAIRSYAVLLHAWPAFAHARILQTNKILVFLRCLLAFVSCICELYFVKACKF 165
Db 121 SPQGLRSYTYLLQGVGFYOKLFNPSPIILFYVRCMLGFGCAWERYMYKSCQSF 180
QY 166 GLHVSRLMLAFVLSTGMFCSSSAFLPSSFCMTYLLIAMTGMWMDKTSIAVLGVAAGAIL 225
Db 181 GIHIGRLWLIFQLPSVGMVFSSTALLPSSFSMYFGCAALAAWQQNYCFALFILTALS 240
QY 226 GWPSAALGLPIAFDILLVMKRWKSPFHSNLSMALILFLVPVVIDSYTYGKLVIAPLN 285
Db 241 GWPFAALIGLPIVLEMLLRDQWKTQVQWTLISGATVAPNTAIDTSYFGKLTFAPLN 300
QY 286 LYNVFTPHGDLXGTEPWYFYLLNGFLNFVAFALALLVPLTSLMEYLLQRFVON-IG 344
Db 301 WYNVFTSHGNIFCTEPLSYIINGFLNFNIIWLLA-LQLPIMLVIDYLVPAKSKSTLN 359
QY 345 HPYVLTAPMYINFIFFIQPHKEERFLFPVYPICLCGAVALSALOKCY----HFVFOR 400
Db 360 FPHYISLAPLYLLVFAQPHKEERFLFPYIPLISLCAITVDVYVYFRFMKSVVFKI 419
QY 401 YRLEHYVTWGNALGTVFLGLLSFSRVALFRGVHGLDLYPEFYRI-AT---DPTTH 456
Db 420 KAGVHYLDHSMFIALVMVSTLLGLSRVYALYRNHAPMDLMLNQLKATPQYDQDV- 478
QY 457 TVPEGRPVNVCWKWYRFPSSFLP-DNKQLOFISFEGQLPKPPAEGPLATRIPTMD 515
```

```

b 479 -----IYNVCIGKDWHRYPGSPFPFAKNFRLFLKGEFRGMLPAYVDEGONATKVQVPY 532
y 516 MUDQNLBPSRVIDISKHYLVLDLTYRETREPKYSGNSKEWISLAYRFLDASSSKL 575
b 533 FNDLNOENHMYFDYDRCDLVDDEGKYTALEPNYSKRSDWMSVMSKSLFLLPEKSHKV 592
y 576 LRAFYVFLSDQTYVYVNTIILKPKAKQ 604
b 593 LRAFYVFLTDNHIQYGFNLK-RKTKR 620

RESULT 9
98GV0
D Q9BGV0 PRELIMINARY; PRT; 198 AA.
C C9BGV0;
T 01-JUN-2001 (TremBLrel. 17, Created)
T 01-JUN-2001 (TremBLrel. 17, Last sequence update)
T 01-DEC-2001 (TremBLrel. 19, Last annotation update)
E Hypochemical 22.0 kDa protein.
S Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
C Cercopitheidae; Macaca.
X NCBI_TaxID=9541;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=frontal cortex;
A Oeada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
A Suzuki Y., Sugano S., Hashimoto K.;
T "Isolation of full-length cDNA clones from macaque brain cDNA
T libraries.";
L Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; AB055306; BAB21931.1; -.
W Hypochemical protein.
Q SEQUENCE 198 AA; 22001 MW; 9276CF0A4D4F3F82 CRC64;

Query Match 30.5%; Score 997; DB 6; Length 198;
Best Local Similarity 95.9%; Pred. No. 5.1e-73;
Matches 198; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

y 1 MASRGARQRLKSGASSGDTAAADKLRELLSRRAGGAHRTLSGNKAGQVWAPGSGT 60
b 1 MASRGARQRLKSGSSGDTAAADKLRELLSRRAGGAHRTLSGNKAGQVWAPGSGT 60
y 61 AFKLLSARLCAALLSNISDCDETNYWEPHYLYIGEGFTWYSPAYAIRSYAYLLH 120
b 61 AFKLLSARLCAALLSNISDCDETNYWEPHYLYIGEGFTWYSPAYAIRSYAYLLH 120
y 121 AWPAAFAHRILOTNKILVYFRLCLLAFVSCICELYFYKAVCKKFGHLVSRMMLAFVL 180
b 121 AWPAAFAHRILOTNKILVYFRLCLLAFVSCICELYFYKAVCKKFGHLVSRMMLAFVL 180
y 181 TGMFCSSSAFLPSSFC 196
b 181 TGMFCSSVVPSPFLIC 196

RESULT 10
18LRHS
D Q8LRHS PRELIMINARY; PRT; 596 AA.
C Q8LRHS;
T 01-OCT-2002 (TremBLrel. 22, Created)
T 01-OCT-2002 (TremBLrel. 22, Last sequence update)
T 01-MAR-2003 (TremBLrel. 23, Last annotation update)
E OSJNBA0016109.26 protein.
N OSJNBA0016109.26
S Oryza sativa (japonica cultivar-group).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
C Ehrhartoideae; Oryzeae; Oryza.
X NCBI_TaxID=39947;
N [1]
P SEQUENCE FROM N.A.

```

```

RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone OSJNBA0016109.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003052; BAB92174.1; -.
DR Gramene; Q8LRHS; -.
SQ SEQUENCE 596 AA; 67812 MW; 839AE12C2DF32292 CRC64;

Query Match 25.9%; Score 849.5; DB 10; Length 596;
Best Local Similarity 34.1%; Pred. No. 1.7e-60;
Matches 217; Conservative 84; Mismatches 249; Indels 87; Gaps 16;

QY 1 MASRGARQRLKSGAS-----SGTAPAAADKLRELLSRRAGGAHRTLSGNKAGQVWAP 56
Db 1 MSLSSARQR-RATAASPTDGGYSKEAKORRRRPSGEEEGIR-----W---44
QY 57 EGSTAFKLLSARLCAALLSNISDCDETNYWEPHYLYIGEGFTWYSPAYAIRSYAY 116
Db 45 --FUPFLALGLLRHMSASSNLHDCDEVFNWEPHFLLYRSFGFTWYSSNFALRSYLY 102
QY 117 LLLHANPAFAHRILOTNKILVYFRLCLLAFVSCICELYFYKAVCKKFGHLVSRMMLAF 176
Db 103 LFIHALVAGPASMIFGEHKRVFVSRIPLGLISTITETVLVVALSRKYGKRLACVLLAM 162
QY 177 LVLTGMFCSSSAFLPSSFCWYTTLIANTGMWMDKTSIAVLGVAAGAILGWPFSAALGLP 236
Db 163 LCLTSGGFFAST-FLPSFSFMYATLSSALFILENVAASVAAAGVILGWPFSLVFLP 221
QY 237 IAFDLLVMKHWKSFHWSMALILFLVPPVVVDSYYGKLVIAPLNIVLVNVTTHGPD 296
Db 222 ITVYSLI-RGSFRVFLSGFLTSNMLLVLSVIADYCYGKWTASVFNLLIKYNVFGGESH 280
QY 297 LYGTPEWVYFLINGFLNFVAFALALLVLPLTSLMEYLLORFHVQNLGHPYMLTLAPMYI 356
Db 281 LYGTGSPFYFNGFNENFAFALLFLGFVP----PARKKYVPL-----LIVSPVYI 332
QY 357 WFIIFFIQHPKEERFLPPVPLICLCAVALSALQKCYHFVORYLEHVTVS-NWLAL 415
Db 333 WLAFMLQAHKEERFLPIYPLICVAAAVIDPDPDFH---DKYSSEQSIAGLRLPLIL 389
QY 416 GTVFLFGLLSFSRVALFRGVHGLDLYPEFYRIATDPTIHTVPEGRFNVVCYGEWYRF 475
Db 390 GFIL---CASHRTFSMLNGCAFIYQHLE-----HHEDTGGSVLCVGSWHRY 438
QY 476 PSSFLP---DNMQLOFIPSEFRGQLPKFPABGPLATRIPTMDNDQNLSEPSRYI--- 528
Db 439 PSSFPPIPSYISEVW---WIDGFRGLLPFPFNETLGTGTTAAPSVDNDKKKASDEQYVFL 495
QY 529 -----DISKCHYLVDLDTMRETREPCKYSSNKEWISL 561
Db 496 APLKYSTVFLITLIRVVTSLHFWLQKDIACSLVVELDLRRYPFR---GSLSTWEYL 552
QY 562 AYRPFLDASRSSKILRAFYVFLSDQTYVYVNTILX 598
Db 553 AALPFLDRELSPALYRSFPFHRWQNNVFLYKLR 589

RESULT 11
Q9FZ49
ID Q9FZ49 PRELIMINARY; PRT; 570 AA.
AC Q9FZ49;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE F611.10 protein (Ser/Thr protein kinase, putative).
GN F611.10 OR AT1G16900.
OS Arabidopsis thaliana (Mouse-ear cress).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```







```

RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Hallsworth K.;
RT "The sequence of C. elegans cosmid T27F7."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Oriol R., Martinez-Duncker I., Chantret I., Mollicone R., Codogno P.;
RT "Common origin and evolution of glycosyltransferases using Dol-P-
monosaccharides as donor substrate."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58762; AAK39304.2; -.
DR EMBL; AJ431373; CAD24083.1; -.
DR WormPep; T27F7.3a; CE30458.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
KW Hypothetical protein; Glycosyltransferase; Transferase.
SQ SEQUENCE 496 AA; 57191 MW; 6A0C53E5346D7F8C CRC64;

Query Match      8.1%; Score 264.5; DB 5; Length 496;
Best Local Similarity 24.4%; Pred. No. 3.9e-13;
Matches 102; Conservative 71; Mismatches 156; Indels 89; Gaps 19;

2Y 62 FKCLLSARLCAALLSNISDCDENWYEPETHLYIYGEFOTWEYSPAYAIRSYAYLLIHA 121
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 14 FRC-LSVYLVASWFPV---DEVYQSAEVAHVLVYGTGHLSEW--RHSLSRFFHPALIA 66

2Y 122 WPAAFHARILQTNKILVYFURCLLAFVSCICELYFYK---AVCKFGFLHVSRRMLAFIV 178
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 67 VIIRKFDILSLSSQFLIYHIPRLAHALLFALADFSFYKICIRCKTKGI-AENSFVYLS 125

2Y 179 LSTGMFCSSAPLPSSPCMYTTLTAMTGWYMDKTSIAVLGVAAGAILG--WPFSALGL- 235
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 126 SWFVFCAPRT-LNSLETSLTIAL-NWPFPEYK-----NFKGTWY-IALGVYL 173

2Y 236 -----PIAFDLLV--MKHRWKS-----PFHWSLMALILFLVPPVVVDSYVYGGKLVIA 281
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 174 TIIIRPTVALIWLFGVYVYHSPVRLIFQLVPLVTLPIVTLIDSWAYGTPTIPL 233

2Y 282 LNVILVNVFTPHGPDLYCTEPWYFLNGFLNFVAALALLVPLTSLMEXYLQRFHVQ 341
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 234 WNFQFNV-VGGGALFGVHFWYIIVSGI-----PAVLTVQMIPIIVGLG----P 280

2Y 342 NLGHFMYTLAPMYIYFIPTIOPHKEERFLFPVYPLICLCGAVALSALQKCYHVFQRY 401
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 281 NIFRPSLLPFPATITYIIVHSLPHKQRFLLPIPLLCIYAGGAFQNLK----- 331

2Y 402 RLEHYVTTSNW---LALGTVFLGLLSFS-----RSVALRFGVHGPLDL 442
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 332 -----WRGSAMVMAINIGIALFTSRVHQVGFPTAPRIMEWRGHRGKLSV 379

```

Search completed: January 13, 2004, 07:03:17  
Job time : 67 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 06:40:49 ; Search time 22 Seconds  
(without alignments)  
1175.088 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGASSGDT.....VNYTLKPKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pcp.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pcp.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCUTS-COMB.pcp.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	131.5	4.0	524	4	US-09-198-452A-52
2	114	3.5	687	3	US-08-834-467-2
3	114	3.5	687	3	US-09-396-177-2
4	113.5	3.5	597	4	US-09-328-352-4703
5	113	3.5	866	4	US-09-134-001C-4930
6	112	3.4	420	4	US-08-795-876-33
7	111	3.4	940	4	US-09-328-352-8165
8	110.5	3.4	423	4	US-08-795-876-38
9	110	3.4	316	4	US-09-549-848B-33
10	110	3.4	538	4	US-09-252-991A-23060
11	109.5	3.3	695	1	US-08-487-886-2
12	109.5	3.3	695	3	US-08-482-855-2
13	109.5	3.3	695	4	US-08-474-986-2
14	108	3.3	638	1	US-08-295-814E-13
15	108	3.3	638	1	US-08-240-783B-2
16	108	3.3	638	3	US-09-084-813-2
17	108	3.3	638	3	US-09-343-361-13
18	108	3.3	638	5	PCT-US92-09662-2
19	107	3.3	390	4	US-09-679-279-5
20	105.5	3.2	436	4	US-08-795-876-2
21	105.5	3.2	692	3	US-07-757-342D-6
22	105.5	3.2	1956	3	US-08-843-417-10
23	105.5	3.2	1956	4	US-09-527-013-10
24	104.5	3.2	422	4	US-09-625-188-12
25	104.5	3.2	3169	2	US-08-477-451-6
26	103.5	3.2	418	4	US-09-252-991A-17598
27	103	3.1	370	3	US-08-513-974B-26

28	103	3.1	370	3	US-08-513-974B-323	Sequence 323, Appl
29	103	3.1	370	3	US-09-172-353-5	Sequence 5, Appli
30	103	3.1	370	3	US-08-776-971-21	Sequence 21, Appl
31	103	3.1	370	3	US-08-776-971-104	Sequence 104, Appl
32	103	3.1	370	4	US-09-799-955-5	Sequence 5, Appli
33	103	3.1	370	4	US-09-461-436B-26	Sequence 26, Appl
34	101.5	3.1	418	3	US-09-030-267-5	Sequence 5, Appli
35	101	3.1	3200	2	US-08-477-451-8	Sequence 8, Appli
36	100.5	3.1	739	4	US-09-134-001C-3586	Sequence 3586, Ap
37	100	3.1	373	4	US-09-622-439-6	Sequence 6, Appli
38	100	3.1	439	4	US-09-107-532A-4168	Sequence 4168, Ap
39	99	3.0	373	4	US-09-622-439-26	Sequence 26, Appl
40	99	3.0	764	3	US-07-741-453A-54	Sequence 54, Appl
41	99	3.0	764	3	US-07-741-453A-60	Sequence 60, Appl
42	99	3.0	764	3	US-07-741-453A-61	Sequence 61, Appl
43	98.5	3.0	533	1	US-09-252-991A-21407	Sequence 21407, A
44	98.5	3.0	3031	1	US-07-689-008-2	Sequence 2, Appli
45	98	3.0	446	2	US-08-672-814D-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1  
US-09-198-452A-52  
; Sequence 52, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Grifffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment thereof and uses thereof, in particular for the diagnosis, prev  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 52  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 1...524  
; OTHER INFORMATION: Xaa=unknown or other  
US-09-198-452A-52

Query Match	4.0%	Score	131.5;	DB	4;	Length	524;
Best Local Similarity	24.1%	Pred. No.	5.3e-05;				
Matches	74;	Conservative	41;	Mismatches	103;	Indels	89;
Gaps	17;						
QY	107	PAYAIRSVAYLLHAWPAAPHARILQTNKILVFYFLRCLLAFVSCICELVFYKAVCKKFG	166				
Db	233	PSSEVGFPASLLIGTLPQNLRLDFRQGLSHLFAISGWHFSLCATLLWM---	LCALLP	289			
QY	167	LHVRMMLAFVLVSTGMFCSSAPLPSSFCMYTLLIAMT-----GWMDKTSIAVLGVAAG	222				
Db	230	LKIKK-ILSETVLT-----SLSCIFPMSLSVMSRWSIVTLLCFSWCFSGCSGLNRLGAG	343				
QY	223	AIL-----GWPFSAALGL---PIAFDLVMKHWKSPF--HW-----	254				
Db	344	FILCSIFSPSPFTVLSTFLATGILLFFPKISFYLP--WTQFLSPFLPIRVIAMT	401				
QY	255	---SLMALILFVPPVVIDSYGKLVIAPIANVLNVNTPHGDLYGTPEWYFVYVINGF	311				
Db	402	LALSIAQLFTVLPIM---QYFGSL---PLEGLYNLIVP-----FTILP---	440				
QY	312	LNFNVAVALALLVPLTS-LMEYLLQRFHQVQNGHPV-----MLTLAPMYIV---	357				
Db	441	---IIVFLIATIIILPCSPITEALIQGF-----LSHPWLNHPNLTLSFAPVPPWMLTLA	493				
QY	358	-FIIFFI 363					
Db	494	SLILFFI 500					

RESULT 2

S-08-834-467-2

Sequence 2, Application US/08834467

Patent No. 608015

GENERAL INFORMATION:

APPLICANT: Albert, Vivian

TITLE OF INVENTION: Glycine Transporter

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dechert Price & Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/834,467

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 314572-109

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 687 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

S-08-834-467-2

Query Match 3.5%; Score 114; DB 3; Length 687;

Best Local Similarity 20.4%; Pred. No. 0.006;

Matches 97; Conservative 50; Mismatches 174; Indels 154; Gaps 20;

Y 84 TNYWEPHYLYG---EGFQWEYSPAYAIRSYAYLLHAWPAAPFARILQTNKILVY 140  
b 302 TFPYVVLTLFVRGVTLEG-----AFDGMWYLTLPQDKILEAKWGDAAASQIFY 351  
Y 141 FLRC-----LLAFVSCICELY-----FYKAVCKKFGHLVS 170  
b 352 SLACAWGGLITWASYNKFNHNCYRDSVIISITNCATSVAGFVIFSLGFMANHLGVDVS 411  
Y 171 RWM-----LAFVLVLTGM-----FCSSSAFLPSSFCWYVTLIAM-----T 205  
b 412 RVADHGPGGLAVAYPEALTLPLSPMLSLFFFWMLILGLGTQFCLETLVTAIVDEGN 471  
Y 206 GWYMDKTSIAVLGVA-AGAILGWPFSAALGLPIAFDILLVMKRWKSPFHWLSMALILFLV 264  
b 472 EWILQKTYVTGLGAVAGFLGLPLTSQAGI---YWLMLMDNYAASF-----SLVVISC 522  
Y 265 PVVVIDSYVYKGLVIAPINIVLVNFTPHGPDLYGTEPWYFYLINGFNFAFALALLV 324  
b 523 IMCVALMIYIGH-----RNVFQDIQMMGLGPPPLFPFQICWRP-----VSPAIIFI 568  
Y 325 LPLTSLMEYLLORFQVNLGHYPWLTLPAPMYIWFITFIQPKBERFLFPVYPPLICLGA 384  
b 569 LVFTVIOYQPIITNHHQ---YFGW-----AAVAGFLMAL-----SSVLCLPIYAMFLCRT 616

QY 385 VALSALQKCHTFVQRVLEHYTTS-NWALGTVFLGLLSFSSRSVALFRYHGPDLILY 443  
Db 617 DGDTLQ-----RLKNATKPSRDW-----GPALE 641  
QY 444 PEFYRIATDPTIHTVPEGRPVNVGVKWEYRFPSSFLLPDNWLOQFIPSEFRGQL 498  
Db 642 HRTGRYA--PTIAPSPED-----GFVEQSLHFDKQAPIVGSNGSRL 682

RESULT 3

US-09-396-177-2

Sequence 2, Application US/09396177

Patent No. 6251617

GENERAL INFORMATION:

APPLICANT: Albert, Vivian

TITLE OF INVENTION: Glycine Transporter

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dechert Price & Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/396,177

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/834,467

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 314572-109

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 687 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-396-177-2

Query Match 3.5%; Score 114; DB 3; Length 687;

Best Local Similarity 20.4%; Pred. No. 0.006;

Matches 97; Conservative 50; Mismatches 174; Indels 154; Gaps 20;

QY 84 TNYWEPHYLYG---EGFQWEYSPAYAIRSYAYLLHAWPAAPFARILQTNKILVY 140  
Db 302 TFPYVVLTLFVRGVTLEG-----AFDGMWYLTLPQDKILEAKWGDAAASQIFY 351  
QY 141 FLRC-----LLAFVSCICELY-----FYKAVCKKFGHLVS 170  
Db 352 SLACAWGGLITWASYNKFNHNCYRDSVIISITNCATSVAGFVIFSLGFMANHLGVDVS 411  
QY 171 RWM-----LAFVLVLTGM-----FCSSSAFLPSSFCWYVTLIAM-----T 205  
Db 412 RVADHGPGGLAVAYPEALTLPLSPMLSLFFFWMLILGLGTQFCLETLVTAIVDEGN 471  
QY 206 GWYMDKTSIAVLGVA-AGAILGWPFSAALGLPIAFDILLVMKRWKSPFHWLSMALILFLV 264  
Db 472 EWILQKTYVTGLGAVAGFLGLPLTSQAGI---YWLMLMDNYAASF-----SLVVISC 522  
QY 265 PVVVIDSYVYKGLVIAPINIVLVNFTPHGPDLYGTEPWYFYLINGFNFAFALALLV 324



TITLE OF INVENTION: STRATEGY TO CLONE DRUGS FOR G PROTEIN  
 TITLE OF INVENTION: COUPLED RECEPTORS  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
 STREET: Clinton Square, P.O. Box 1051  
 CITY: Rochester  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 14603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/795,876  
 FILING DATE:  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BRAMAN, SUSAN J.  
 REGISTRATION NUMBER: 34,103  
 REFERENCE/DOCKET NUMBER: 19603/1280  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 716-263-1636  
 TELEFAX: 716-263-1600  
 INFORMATION FOR SEQ ID NO: 33:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 420 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 S-08-795-876-33

Query Match 3.4%; Score 112; DB 4; Length 420;  
 Best Local Similarity 21.5%; Pred. No. 0.0044;  
 Matches 90; Conservative 56; Mismatches 143; Indels 130; Gaps 25;  
 Y 70 LCAALLSNISDCDETFN---YWEPP-----THVLI---YGEQFTWEYSPAY-----A 110  
 b 24 LCGVVSDYKDDDDKDLATLFWEDBERKESGLTEYRLVSNKSSPLQKLPATISEDASG 83  
 Y 111 IRSYAVLLHAWPAAFHA-----RILOTN--KILVYFLRCLLAFVSCICELVPEYKA 160  
 b 84 YLGNILRLVLIWFISILAITGNIIVILVITTSQYKLTVPFLMCLNLAFLAD-LC-IGIYLL 141  
 Y 161 VCKKFGHLVSRMMLAPLV-LSTGMFCSSAP---LPSSFCMYT-TLIAMTGWY----- 208  
 b 142 LIASVDIHTKSQVHNAIDWQTGAGDAAGFTVPFASLSVTLTATLERWHTITHAMQ 201  
 Y 209 MDKTSIAVLGVAAGAILGWPRSA-----LGLPIAFD-----LLVMKHR 247  
 b 202 LD-CRKQLRHAASVMVGWIFAFALPFIPIGSISSMKVSCLEPMDIDSPLSQLYVM--- 257  
 Y 248 WKSFFHWSLMALILFLVPVVVDSYVYKGLVPIAPNLVLYN-----VFTPHGP 295  
 b 258 -----SLVLNV-LAFVVICGVTHIYLTVRNPVIVSSSDTRIARMAAMLIPT--- 305  
 Y 296 DLYGTEPWYFYLINGFLNFVAFALALVPL--TSLMEYLLQRPV-VQNLGHEPYWLTLA 352  
 b 306 DFLCMAPISFFAIS-----ASLKVPLITVSKAKILLVLFHPINSCANFLYAI- 353  
 Y 353 PMYINFIPIQPHKEERLFFVPLICLCGAVALSALQKCYHVFQRYLHYTVTSN 411  
 b 354 -----FTKNFRDRDFI-----LLSKCG-----CYEMQAIQYRTSTSTVHN 389

RESULT 7  
 S-09-328-352-8165  
 Sequence 8165, Application US/09328352  
 Patent No. 6562958  
 GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: GTC99-03PA  
 CURRENT APPLICATION NUMBER: US/09/328,352  
 CURRENT FILING DATE: 1999-06-04  
 NUMBER OF SEQ ID NOS: 8252  
 SEQ ID NO 8165  
 LENGTH: 940  
 TYPE: PRT  
 ORGANISM: Acinetobacter baumannii  
 US-09-328-352-8165

Query Match 3.4%; Score 111; DB 4; Length 940;  
 Best Local Similarity 20.6%; Pred. No. 0.021;  
 Matches 81; Conservative 57; Mismatches 154; Indels 102; Gaps 21;  
 QY 87 YWETHYLIYGEQFTWEYSPAYAIRSYAYLLHAWPAAFHARILQTNKILVYFLRCLL 146  
 DB 499 HMTAIGWVAQGTALIVMGVTYRLSRYIGVIL-----VLLSLALFYQWANE 547  
 QY 147 AFVSCICELYFYKAVCKKFGHLVSRMMLAPLVLTG-----MFCSSAFPLPSFCMYTTLI 202  
 DB 548 EFTLSTSIY-----ALAQFISAFYLLQYNSKEQRYFSASMFSGIFLC----- 590  
 QY 203 AMTGWYMDKTSIAVLGVAAGAILG-----WPFSA-----ALGLPIAFDLIV---MKHR 247  
 DB 591 -----LGNVAGAVAGVEIMAWHHALSPLYLMFAIALIAFSAIVHYKLRVQ 636  
 QY 248 WKSFFHWSLMALILFLV-----PVVV-----IDSVYVYKGLVPIAPNLVLYNVPFPHGPD 296  
 DB 637 WQS-LQLLILSLILLVLGGEAFASQVFTLTKWDSLQQTFLVS---TILSGLFMAQVQ 693  
 QY 297 -----LYGTEPWYFYLINGFLNFVAFALALVPLT-SLMEYLLQRPVQNLGH--- 345  
 DB 694 SSHLGYVYKVMAGLSNLALAVGVAIFPKMPVIALAFVVPVYSLWAY---KAHKTTLLHQI 750  
 QY 346 PW-LTLAPWYINFIPIQPHKEERFLF-PVYPLICLCGAVALSALQKCYHVFQRYL 403  
 DB 751 PWCLSL-----IWLIVVSDRHSAEVLYFVPLFNLITDFFSILVFAGLL-----FIYCHAP 802  
 QY 404 EHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYH 437  
 DB 803 DQ-DKSLWTFKITTILVGLLVFSSVVV--RGLH 833

RESULT 8  
 US-08-795-876-38  
 Sequence 38, Application US/08795876  
 Patent No. 6403305  
 GENERAL INFORMATION:  
 APPLICANT: Gershengorn, Marvin C.  
 APPLICANT: Geras-Raaka, Elizabeth  
 APPLICANT: Nussenzweig, Daniel R.  
 TITLE OF INVENTION: STRATEGY TO CLONE DRUGS FOR G PROTEIN  
 TITLE OF INVENTION: COUPLED RECEPTORS  
 NUMBER OF SEQUENCES: 44  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
 STREET: Clinton Square, P.O. Box 1051  
 CITY: Rochester  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 14603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/795,876  
 FILING DATE:  
 CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: BRAMAN, SUSAN J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 19603/1280  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 716-263-1636  
TELEFAX: 716-263-1600  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-795-876-38

Query Match 3.4%; Score 110.5; DB 4; Length 423;  
Best Local Similarity 21.3%; Pred. No. 0.0065;  
Matches 90; Conservative 56; Mismatches 143; Indels 133; Gaps 25;

QY 70 LCAALLSNISCDTFN-----YNEP-----THYLI--YGEGFQWEYSAY-----109  
DB 24 LCOGVSDYKDDDDKLDATLWPFWEDEEKNESGLTEYRLVNSINKSPLOKQLPAFISED 83  
QY 110 --AIRSYAYLLHAWPAFHA-----RILOTN--KILVVFELRCLLAFVSCICELYF 157  
DB 84 ASGYLGYNILRVLIWIFISILAITGNIIVLITTSQYKLVPRFLMCLNLAFAAD-LC-IGI 141  
QY 158 YKAVCKEGLHVSRRMLAFIV-LSTGMFCSSAF---LPSFCMYT-TLIAMTGWY-----208  
DB 142 YLLIASVDIHTKQYHNYAIDWOTGAGCDAAGFTVFASELSYTYLTALTIERMHTITH 201  
QY 209 ---MDKTSIAVLGVAAGAILGWPFSAA-----LGLPIAPD-----LLVM 244  
DB 202 AMQLD-CKVQLRHAASVMWGWIFAPAAALPPIGISSYMKVSIICLPMDIDSPLSQLYVM 260  
QY 245 KIRWKSFFHNSLMALILFLVPPVVIDSYHYGKLVAPLNVLYN-----VFTP 292  
DB 261 -----SLLVNV-LAFWVCGCYHIHLYLVNPNIVSSSDTRIAKRMAMLIPT- 308  
QY 293 HGPDLVGTPEPYFVYLINGFLNFVAFALALVLPL--TSLMEYLQRFH-VONLGHYPYL 349  
DB 309 ---DPLWAPISFAIS-----ASLKVPLIVTSKAKILLVLFHPINSCANPFY 354  
QY 350 TLAEMYIPIFFTOPHKEERFLPPVPLICLCAVALSALQCKYHFVORYLEHYTVT 409  
DB 355 AI-----FTKNFRDRDFI-----LISKCG-----CYEMQAQIYRTETSTV 390  
QY 410 SN 411  
DB 391 HN 392

Query Match 3.4%; Score 110.5; DB 4; Length 423;  
Best Local Similarity 21.3%; Pred. No. 0.0065;  
Matches 90; Conservative 56; Mismatches 143; Indels 133; Gaps 25;

QY 70 LCAALLSNISCDTFN-----YNEP-----THYLI--YGEGFQWEYSAY-----109  
DB 24 LCOGVSDYKDDDDKLDATLWPFWEDEEKNESGLTEYRLVNSINKSPLOKQLPAFISED 83  
QY 110 --AIRSYAYLLHAWPAFHA-----RILOTN--KILVVFELRCLLAFVSCICELYF 157  
DB 84 ASGYLGYNILRVLIWIFISILAITGNIIVLITTSQYKLVPRFLMCLNLAFAAD-LC-IGI 141  
QY 158 YKAVCKEGLHVSRRMLAFIV-LSTGMFCSSAF---LPSFCMYT-TLIAMTGWY-----208  
DB 142 YLLIASVDIHTKQYHNYAIDWOTGAGCDAAGFTVFASELSYTYLTALTIERMHTITH 201  
QY 209 ---MDKTSIAVLGVAAGAILGWPFSAA-----LGLPIAPD-----LLVM 244  
DB 202 AMQLD-CKVQLRHAASVMWGWIFAPAAALPPIGISSYMKVSIICLPMDIDSPLSQLYVM 260  
QY 245 KIRWKSFFHNSLMALILFLVPPVVIDSYHYGKLVAPLNVLYN-----VFTP 292  
DB 261 -----SLLVNV-LAFWVCGCYHIHLYLVNPNIVSSSDTRIAKRMAMLIPT- 308  
QY 293 HGPDLVGTPEPYFVYLINGFLNFVAFALALVLPL--TSLMEYLQRFH-VONLGHYPYL 349  
DB 309 ---DPLWAPISFAIS-----ASLKVPLIVTSKAKILLVLFHPINSCANPFY 354  
QY 350 TLAEMYIPIFFTOPHKEERFLPPVPLICLCAVALSALQCKYHFVORYLEHYTVT 409  
DB 355 AI-----FTKNFRDRDFI-----LISKCG-----CYEMQAQIYRTETSTV 390  
QY 410 SN 411  
DB 391 HN 392

SEQ ID NO 33  
LENGTH: 316  
TYPE: PRT  
ORGANISM: Synechocystis sp  
US-09-549-848B-33

Query Match 3.4%; Score 110; DB 4; Length 316;  
Best Local Similarity 22.6%; Pred. No. 0.0046;  
Matches 61; Conservative 43; Mismatches 86; Indels 86; Gaps 13;

QY 173 MLAFVLVSTGMFCSSAFLPSSFCMTTLIAMTGWYMDKTS-LAVLGVAAGAI---LGWP 228  
DB 110 VLSPALMATFNVLSGLALSIVFY--MLVTHWLKRHTAQNIVIGGGAAGSIPPLVGM- 166  
QY 229 FSAALGLPIAFDILLVMKRWKSFHSLMALILFLVP-----VVVIDSYHYGKLVAP 281  
DB 167 -AAVTG-----DL-----SWTPWVLFALIFLWTPPHFWALALMIKDDY-----AQ 205  
QY 282 LNVLYNVVTFPHGPDLYGTEPYFVYLINGFLNFVAFALALVLPLTSLMEYLLORFHVQ 341  
DB 206 VNVPLPVIA--GEETVSIWIY-----SLLVVPFSLLVTPHQLGI- 247  
QY 342 NLGHPYLTLPAPYIWIPIFFIQPHKEER-----FLFPVPLICLCAVALSAL 390  
DB 248 -----LYLAIAIILGGQFLVKAQKQAPGDRDLARGLFXFSFYLMLLCLAWIDSL 300  
QY 391 QKCYHFVORYLEHYTVTSNWLALGTVEL 420  
DB 301 PVTHQLVAQ-----MGTLLL 315

RESULT 10  
US-09-252-991A-23060  
Sequence 23060, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 23060  
LENGTH: 538  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23060

Query Match 3.4%; Score 110; DB 4; Length 538;  
Best Local Similarity 22.9%; Pred. No. 0.011;  
Matches 75; Conservative 39; Mismatches 96; Indels 118; Gaps 20;

QY 167 LHVSRRMLAF---LVLSGMFC---SSSAPLSSPCMYTTLIAMTGWYMDKTSIAVLGVA 220  
DB 123 LHTNAVIFAFGGCAIFATSYVYVQRTSQARLISD-----TLAFTFWGQ-----AVI 170  
QY 221 AGAILGWP--FS-----AALGLPIAFDILLVMKRWKSFHSLMALILF-----LVPV 267  
DB 171 VGAVLTLPQGFTTSKEYAELEWPLALLAIV-----WITYAIVFPGTIVKRVKXHI 221  
QY 268 VIDSYYGK--LVIAPLNV-----LYNVTPHGPDLVGTPEPY-----FYLI 308  
DB 222 YVGNWFYGAFLVLTAMLHVNHMSLPVSWFKSYSAYS--GATDAMVQWVYGHNAVGFELT 279  
QY 309 NGFLNFVAFALALLVLPLTSLMEYLQRFHVQNLGHPYWLTLAPMYTWIFFIOPHKE 368  
DB 280 TGFLGMWYFVPKQARPYVS---YRLSIHV-----FW-ALISLIYW-----AGPH-- 321

y 369 ERFLPVPVPLICLGAVALSALQKCYHFVORYLEHYVTTSNWL-ALGTVFLLGLISFS 427  
 b 322 -----HL-----HYTLPDWAQSLGVMVMSILLIAPS 347  
 y 428 RVALFRGHGDLDPYEFYRIATDPTI 455  
 b 348 WGGMI-----NGWMTLSGAWHKLRTDPII 371  
 RESULT 11  
 S-08-487-886-2  
 Sequence 2, Application US/08487886  
 Patent No. 5744448  
 GENERAL INFORMATION:  
 APPLICANT: Kelton, Christie Ann  
 APPLICANT: Schweickhardt, Rene Lynn  
 APPLICANT: Cheng, Shirley Vui Yen  
 APPLICANT: Nugent, No. 5744448een Patrice  
 TITLE OF INVENTION: Human Pollicle Stimulating  
 TITLE OF INVENTION: Hormone Receptor  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSES: Stephan P. Williams,  
 ADDRESSES: Ares-Serono, Inc.  
 STREET: Exchange place, 37th floor  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density  
 COMPUTER: IBM PS/2, model 55 SX  
 OPERATING SYSTEM: MS-DOS version 4.0  
 SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/487,886  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/670,085  
 FILING DATE: 15-MAR-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Williams, Stephan P.  
 REGISTRATION NUMBER: 28546  
 REFERENCE/DOCKET NUMBER: US/252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 723-1300  
 TELEFAX: (617) 723-8923  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 695  
 TYPE: Amino acid  
 TOPOLOGY: Linear  
 MOLECULE TYPE: protein  
 FEATURE:  
 NAME/KEY: signal sequence  
 LOCATION: -17 to -1  
 IDENTIFICATION METHOD: hydrophobic  
 FEATURE:  
 NAME/KEY: putative amino-terminal extracellular domain  
 LOCATION: 1 to 349  
 IDENTIFICATION METHOD: similarity with other  
 IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular  
 IDENTIFICATION METHOD: domains, hydrophilic  
 FEATURE:  
 NAME/KEY: transmembrane domain  
 LOCATION: 350 to 613  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains  
 FEATURE:  
 NAME/KEY: putative transmembrane region I  
 LOCATION: 350 to 370  
 IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative transmembrane region II  
 LOCATION: 382 to 404  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative transmembrane region III  
 LOCATION: 427 to 448  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative transmembrane region IV  
 LOCATION: 469 to 491  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative transmembrane region V  
 LOCATION: 512 to 533  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative transmembrane region VI  
 LOCATION: 557 to 580  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative transmembrane region VII  
 LOCATION: 592 to 613  
 IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:  
 NAME/KEY: putative carboxy-terminal intracellular  
 NAME/KEY: domain  
 LOCATION: 614 to 678  
 US-08-487-886-2  
 Query Match 3.3%; Score 109.5; DB 1; Length 695;  
 Best Local Similarity 22.2%; Pred. No. 0.018;  
 Matches 81; Conservative 48; Mismatches 125; Indels 111; Gaps 21;  
 QY 105 YSPAVAIRSYAVLLHAWPAAFHA-----RIQTN--KILVVFELCLLAFVSCICE 154  
 Db 353 FNPCEIDMGYNILRVLIWFISILAITGNIIIVLITTSQYKLTVPRLMCNLAFAD-LC- 410  
 QY 155 LYFYKAVCKKFGHVSRRMLAFV-LSTGMFCSSAF---LPSSFCWYT-TLIAMTQWY- 208  
 Db 411 IGIYLLIASVDIHTKSOYHNYAIDWQTGAGCDAAGPFTVFASLSVYTLTALTLEWH 470  
 QY 209 -----MDKTSIAVLGVAAGAILGWPFSAA-----LGLPIAFD-----L 241  
 Db 471 ITHAQQLD-CKVQLRHAAVVMGMWIFAFAPALPEIFGISYKVKVSIPLMDISPLSQL 529  
 QY 242 LYMKHRWKSFFHWSLMALILFLVPVWIDSYYYGKVIAPINIVLYN-----V 289  
 Db 530 YVM-----SLVLNV-LAFWVICGCIYHILTVRNPNIVSSSSDTRIARKMAMLI 578  
 QY 290 FTFPHGPDLYGTPEWYFYLLNGFLNFVAFALALVPLI--TSLMEYLLQRFH-VQNLGHP 346  
 Db 579 FT-----DFLCWAPISFFAIS-----ASKVPLITVSKAKILLVLFHPINSCANP 623  
 QY 347 YMLTLPAPYIWFIFFIQPHKEERFLFPVYVPLICLGAVALSALQKCYHFVORYLEHY 406  
 Db 624 FLYAI-----FTKNPRDFEI-----LLSKCG-----CYEMQAQIYRTETS 659  
 QY 407 TVTSN 411

Db 660 STVHN 664

RESULT 12

US-08-482-855-2

Sequence 2, Application US/08482855

Patent No. 6121016

GENERAL INFORMATION:

APPLICANT: Kelton, Christie Ann  
 APPLICANT: Schweickhardt, Rene Lynn  
 APPLICANT: Cheng, Shirley Vui Yen  
 APPLICANT: Nugent, No. 6121016seen Patrice  
 TITLE OF INVENTION: Human Follicle Stimulating  
 NUMBER OF INVENTIONS: Hormone Receptor

CORRESPONDENCE ADDRESSES: 2

ADDRESSEE: Stephan P. Williams,

ADDRESSEE: Ares-Serono, Inc.

STREET: Exchange Place, 37th floor

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density

COMPUTER: IBM PS/2, model 55 SX

OPERATING SYSTEM: MS-DOS version 4.0

SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/482,855

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/670,085

FILING DATE: 15-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Stephan P.

REGISTRATION NUMBER: 28546

REFERENCE/DOCKET NUMBER: US/252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 723-1300

TELEFAX: (617) 723-8923

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 695

TYPE: Amino acid

TOPOLOGY: Linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: signal sequence

LOCATION: -17 to -1

IDENTIFICATION METHOD: hydrophobic

FEATURE:

NAME/KEY: putative amino-terminal extracellular domain

LOCATION: 1 to 349

IDENTIFICATION METHOD: similarity with other

IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular

IDENTIFICATION METHOD: domains, hydrophilic

FEATURE:

NAME/KEY: transmembrane domain

LOCATION: 350 to 613

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains

FEATURE:

NAME/KEY: putative transmembrane region I

LOCATION: 350 to 370

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region II

LOCATION: 382 to 404

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

IDENTIFICATION METHOD: similarity to other G  
 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,  
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length  
 FEATURE:

NAME/KEY: putative transmembrane region III

LOCATION: 427 to 448

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region IV

LOCATION: 469 to 491

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region V

LOCATION: 512 to 533

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region VI

LOCATION: 557 to 580

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region VII

LOCATION: 592 to 613

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative carboxy-terminal intracellular

NAME/KEY: domain

LOCATION: 614 to 678

US-08-482-855-2

Query Match 3.3%; Score 109.5; DB 3; Length 695;

Best Local Similarity 22.2%; Pred. No. 0.018;

Matches 81; Conservative 48; Mismatches 125; Indels 111; Gaps 21;

QY 105 YSPAYAIRSYAYLLHAWPAAFHA-----RILOTN--KILVFYFLRCLLAFVSCICE 154

Db 353 PNPCEIDMGYNILRVLIWFI-SILAITGNIIVILVLTTSQYKLTVPREFLNCNLAFAAD-LC- 410

QY 155 LYFYKAVCKKFGHLVSRMLAFIV-LSTGMFCSSAP----LPSSFCMYT-TLIAMTGWY- 208

Db 411 IGIYLLLIASVDIHTKSQYHNYAIDWQTGAGCDAGFTTVFASLSVYTLTATILRWET 470

QY 209 -----MDKTSIAVLGVAAGAILGWPFSA-----LGLPIAFD-----L 241

Db 471 ITHAQILD-CKVQLRHASVWVGWIFAFAPALFPFIFGISSYMKVSCICLPMDIDSPSQL 529

QY 242 LVMKRWKSPFHHSLMALILFLVFPVVVIDSYYGKLVIAPLNIVLYN-----V 289

Db 530 YVM-----SLLVLNV-LAFWICGCIHYILTVRNPNIVSSSSDTRIAKRWAMLI 578

QY 290 FTPHGPDLYGTEPWYFYLLINGFLNENVPALALLPL--TSLMEYLLQRFH-VOMLGH 346

Db 579 FT----DFLCWAPISFFAIS-----ASLKVPILTVSKAKILLVLFHPINSCAP 623

QY 347 YWLTAPMYIWFIFFIQPHKEERFLFPVYPLICLCAVALSALQKCYHFVFORYLEHY 406

Db 624 FLYAI-----FTKNFRDROFFI-----LLSKCG-----CYEMQAQIVRTETS 659

QY 407 TVTSN 411

Db 660 STVHN 664

RESULT 13



3-08-474-986-2

AL INFORMATION:  
 APPLICANT: Kelton, Christie Ann  
               Schweickhardt, Rene Lynn  
               Cheng, Shirley Vui Yen  
               Nugent, No. 637271leen Patrice  
 TITLE OF INVENTION: Human Follicle Stimulating  
                           Hormone Receptor

NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stephan P. Williams,  
Ares-Serono, Inc.  
STREET: Exchange Place, 37th floor

CITY: BOSTON  
STREET: EXCHANGE PLACE, 3700 11001  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density  
COMPUTER: IBM PS/2, model 55 SX  
OPERATING SYSTEM: MS-DOS version 4.0  
SOFTWARE: VAX/VMS Massill via Kermit to IBM MS-DOS  
CURRENT APPLICATION DATA:

SEQUEN  
5-08-474-986-2

Query Match 3.3%; Score 109.5; DB 4; Length 695;  
Best Local Similarity 22.2%; Pred. No. 0.018;  
Matches 81; Conservative 48; Mismatches 125; Indels 111; Gaps 21;

[illegible]

**RESULT 14**

```

US-08-295-814E-13
/ Sequence 13, Application US/08295814E
/ Patent No. 5658786
/ GENERAL INFORMATION:
/ APPLICANT: Smith, Kelli E.
/ APPLICANT: Borden, Laurence A.
/ APPLICANT: Hartig, Paul R.
/ APPLICANT: Weishank, Richard L.
/ TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA
/ TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
/ NUMBER OF SEQUENCES: 35
/ CORRESPONDENCE ADDRESS:
/ ADDRESSSEE: Cooper & Dunham LLP
/ STREET: 1185 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/295,814E
/ FILING DATE: DECEMBER 19, 1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John
/ REGISTRATION NUMBER: 28,678
/ REFERENCE/DOCKET NUMBER: 40558-B-PCT-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-278-0400
/ TELEFAX: 212-391-0525
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 638 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-08-295-814E-13

```

Query Match	3.3%;	Score 108;	DB 1;	Length 638;		
Best Local Similarity	20.4%;	Pred. No. 0.023;				
Matches	75;	Conservative	144;	Indels 104; Gaps 15;		
QY	84	TENWEPEHYLLYG---	EGFQTWEYSPVAIRSYAYLLLHAWPAAFHARILQTNKILVYF	140		
DB	253	TFPYVVLTLFVRGVTLG-----	AFTGIMYYLTPKWDKILEAKWGDAAASQIFY	302		
QY	141	FLRC-----	LLAFVSCICELY-----	FYKAVCKPGLHVS	170	
DB	303	SLGCAGWGGLITMASYNKFNHNCYRDSVIISITNCATSVYAGVFISILGFMANHLGLVDVS	362			
QY	171	RNM-----	LAFLVLSGTM-----	FCSSSAELPSSFCMYITLIAM-----	T 205	
DB	363	RVADHGGPGLFAVYPEALTLPLSPLSLFFPMLILLGLGTQFCLELFTVAIVDEWGN	422			
QY	206	GYWMDKTSIAVLGYA-AGAILGPPSPSAAALGLPIAFDLLVMKHKWKSFFHWSLMALILFLV	264			
DB	423	EWILQKTYVTILGVAVAGFLGLTSLQAGI---	YWLMLDMNYAASF-----	SLVVIS	473	
QY	265	PVVVIDSYYYGKLVIAPLNIVLVNFTPHGSDLYGTERPWYFYLLINGLFNPNVAFALALLV	324			
DB	474	IMCVSIMIYIGH-----	RNYFODIQMMLGPPPLFFQICRWFSPTIIF---	FI	519	
QY	325	LPLTSLMEVLLQRHVQNLGHYPWYLAPWYIWFILFTFIQPHKEERFLPVVPYPLICIGA	384			
DB	520	LIFTVIOVRPITYNHQO---	YPGW---	AVAIGFLNAL-----	SSVICIPLDALFQLCRT	567
QY	385	VALSALQK	392			

Db 568 DGDILLOR 575

Db 520 LIETVIOYRPIITYNHQ---YPCW-----AVAIGFLMAL-----SSVICIPLVAFQLCRT 567

QY 385 VALSALOK 392

QY 385 VALSALOK 392

Db 568 DGDILLOR 575

Db 568 DGDILLOR 575

Search completed: January 13, 2004, 07:03:43  
Job time : 26 secs

RESULT 15  
US-08-240-783B-2  
; Sequence 2, Application US/08240783B  
; Patent No. 5756348  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Kelli  
; APPLICANT: Borden, Laurence A.  
; APPLICANT: Branche, Theresa  
; APPLICANT: Hartig, Paul R.  
; APPLICANT: Weinschenk, Richard L.  
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/240.783B  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 638 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-240-783B-2

Query Match 3.3%; Score 108; DB 1; Length 638;  
Best Local Similarity 20.4%; Pred.No.0.023;  
Matches 75; Conservative 45; Mismatches 144; Indels 104; Gaps 15;  
QY 84 TFNYWEPHYLIYG---EGFQWYSPAYAIRSYAYLLHAWPAAPFARILQTNKILVY 140  
Db 253 TFPYVVLITLVRGVTEG-----AFTGIMVLTTPKWDKILEAKVWGDAASQIFY 302  
QY 141 FLRC-----LLAFVSGICELY-----FYKAVCKKFGHLVS 170  
Db 303 SLGCAGWGGLITWASYNKPHNNCYRDSVITNCATSVIAGVTFISILGFMANHLGVDVS 362  
QY 171 RMW-----LAFVLSTGM-----FCSSAFLPSPFCMYTTLIAM-----T 205  
Db 363 RVADHGPGGLAFVAYPEALTLLPISLWSLLFFFMILLGLGTQCLLETIVTAIVDEVGN 422  
QY 206 GWYDKTSTAVLGVA-AGAILGWPSAALGLPIADLLVMKRWKSPFHWLMAILFLV 264  
Db 423 EWILQKTYTTLGVAVAGFLGPIPTSOAGI---YWLMDNDNYAASF-----SLVVISC 473  
QY 265 PVVVIDSVYVYKGLVIAPLNVLNVFTPHGPDLYGTEPWYFYLLINGFLNFNFAFALLV 324  
Db 474 IMCVSIMYIYGH-----RNVFQDIQMWLGFPFPLFFQICWRVFSPTIIF---FI 519  
QY 325 LPLTSLMEYLLQRFHVQNLGHFWTLAPMWIWIIFFIQPHKEERFLFPVYPLICLGA 384

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 13, 2004, 07:03:24 ; Search time 53 Seconds  
(without alignments)  
2323.065 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGASSGDT.....VNYTLKPRKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	3274	100.0	611	11	US-09-922-225A-2
2	1797	54.9	339	11	US-09-922-225A-8
3	1489	5	45.5	331	US-09-922-225A-6
4	1395	42.6	258	15	US-10-050-704-121
5	1358	41.5	623	11	US-09-922-225A-14
6	1013.5	31.0	255	11	US-09-922-225A-4
7	986	30.1	653	11	US-09-922-225A-13
8	973	29.7	577	11	US-09-922-225A-11
9	973	29.7	577	11	US-09-922-225A-15
10	812	24.8	570	11	US-09-922-225A-12
11	209.5	6.4	554	12	US-10-231-265-240
12	130	4.0	433	16	US-10-080-170-547
13	126	3.8	584	12	US-10-032-585-7898
14	117	3.6	440	16	US-10-080-170-164
15	111.5	3.4	458	9	US-09-862-767A-9

16	111.5	3.4	899	12	US-10-318-306A-48	Sequence 48, Appl
17	111.5	3.4	899	12	US-10-319-236A-48	Sequence 48, Appl
18	111.5	3.4	1210	9	US-09-922-217-692	Sequence 692, App
19	111.5	3.4	1210	10	US-09-833-263-692	Sequence 692, App
20	111.5	3.4	1210	14	US-10-035-380-692	Sequence 692, App
21	111.5	3.4	1548	9	US-09-922-217-1095	Sequence 1095, Ap
22	111.5	3.4	1548	14	US-10-025-380-1095	Sequence 1095, Ap
23	111.5	3.4	460	9	US-09-741-669-464	Sequence 464, App
24	110.5	3.4	695	12	US-10-349-838A-22	Sequence 22, Appl
25	110	3.4	314	12	US-10-017-161-606	Sequence 38, Appl
26	110	3.4	314	12	US-09-908-006A-38	Sequence 38, Appl
27	110	3.4	314	12	US-10-292-798-532	Sequence 532, App
28	110	3.4	216	12	US-10-349-508-33	Sequence 33, Appl
29	110	3.4	755	15	US-10-156-761-12016	Sequence 12016, A
30	109.5	3.3	215	10	US-09-747-155-421	Sequence 421, App
31	109.5	3.3	604	12	US-10-292-798-1170	Sequence 1170, Ap
32	109.5	3.3	695	10	US-09-804-626-8	Sequence 8, Appli
33	109.5	3.3	695	12	US-10-349-838A-2	Sequence 2, Appli
34	109.5	3.3	695	12	US-10-349-838A-26	Sequence 26, Appl
35	109.5	3.3	695	12	US-10-349-838A-28	Sequence 28, Appl
36	109.5	3.3	695	12	US-10-349-838A-30	Sequence 30, Appl
37	109.5	3.3	695	12	US-10-349-838A-34	Sequence 34, Appl
38	109.5	3.3	695	15	US-10-225-567A-122	Sequence 122, App
39	109.5	3.3	695	15	US-10-207-655-65	Sequence 65, Appl
40	108.5	3.3	406	12	US-10-091-007-12	Sequence 12, Appl
41	108	3.3	317	9	US-09-815-242-13638	Sequence 13638, A
42	108	3.3	633	12	US-10-205-219-30	Sequence 30, Appl
43	108	3.3	638	12	US-09-845-908-13	Sequence 13, Appl
44	107.5	3.3	695	12	US-10-349-838A-24	Sequence 24, Appl
45	107.5	3.3	695	12	US-10-349-838A-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-09-922-225A-2  
; Sequence 2, Application US/09922225A  
; Publication No. US20030104385A1  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen A.  
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
; TITLE OF INVENTION: Associated with Bipolar Disorder  
; FILE REFERENCE: P-EA 4672  
; CURRENT APPLICATION NUMBER: US/09/922,225A  
; CURRENT FILING DATE: 2003-01-14  
; NUMBER OF SEQ ID NOS: 117  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 611  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-225A-2

Query Match 100.0%; Score 3274; DB 11; Length 611;  
Best Local Similarity 100.0%; Pred No. 3.9e-309;  
Matches 611; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASRGARQLKSGASSGDTAPAADKLRELLSGREAGAEHRTLSGNKAGQVWPEGST	60
DB	1	MASRGARQLKSGASSGDTAPAADKLRELLSGREAGAEHRTLSGNKAGQVWPEGST	60
QY	61	AFKLLSARLCALLSNISDCDTEFNWEPHTLYLYGEGFQTWESPAVARSAYALLH	120
DB	61	AFKLLSARLCALLSNISDCDTEFNWEPHTLYLYGEGFQTWESPAVARSAYALLH	120
QY	121	AWPAAFHARILQTNKILVFLRCLAFVSCICELFYKAVCKKFGHVSRRMMLAFVLVS	180
DB	121	AWPAAFHARILQTNKILVFLRCLAFVSCICELFYKAVCKKFGHVSRRMMLAFVLVS	180
QY	181	TGMFCSSSAFLPSSFCWYTTLIAMTGMVMDKTSIAVLGVAAGAILGWPFSALGLPIAFD	240
DB	181	TGMFCSSSAFLPSSFCWYTTLIAMTGMVMDKTSIAVLGVAAGAILGWPFSALGLPIAFD	240

241 LLVGMHRKWSFFHSLMALILFLVPVVVVIDSYTGKVIAPLNIIVLVNFTPHGPDLYGT 300  
241 LLVGMHRKWSFFHSLMALILFLVPVVVVIDSYTGKVIAPLNIIVLVNFTPHGPDLYGT 300  
301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVNGLGHPYWLTLAPMYWFI 360  
301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVNGLGHPYWLTLAPMYWFI 360  
361 FTIQPKHERFLFPVYPLICLCGAVALSALOKYHFVQRYRLEHYTVTSNWLALGTVEL 420  
361 FTIQPKHERFLFPVYPLICLCGAVALSALOKYHFVQRYRLEHYTVTSNWLALGTVEL 420  
421 FGLLSFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWYRFFSSFL 480  
421 FGLLSFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWYRFFSSFL 480  
481 LPDNWLOQIPSEFRGOLPKPFAEGPLATRIPTDMDNDQNLLEEPSRIDISKHYLVLDL 540  
481 LPDNWLOQIPSEFRGOLPKPFAEGPLATRIPTDMDNDQNLLEEPSRIDISKHYLVLDL 540  
541 TWRETPREPKYSNKEEWSLAYRPFLLDASRSSKLLRAFYVFFLSDQYTVVNYTILKPR 600  
541 TWRETPREPKYSNKEEWSLAYRPFLLDASRSSKLLRAFYVFFLSDQYTVVNYTILKPR 600  
601 KAKQIRKXSGG 611  
601 KAKQIRKXSGG 611

RESULT 2  
S-09-922-225A-8  
Sequence 8, Application US/09922225A  
Publication No. US20030104385A1  
GENERAL INFORMATION:  
APPLICANT: Evans, Glen A.  
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
FILE REFERENCE: P-EA 4672  
CURRENT APPLICATION NUMBER: US/09/922,225A  
CURRENT FILING DATE: 2003-01-14  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 339  
TYPE: PRT  
ORGANISM: Homo sapiens  
IS-09-922-225A-8

Query Match 54.9%; Score 1797; DB 11; Length 339;  
Best Local Similarity 100.0%; Pred. No. 6.4e-166;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MASRGARQLKSGASGDTAPAADKRLGSLRAGGAEHRTLSGNKAGQWAPGEGST 60  
1 MASRGARQLKSGASGDTAPAADKRLGSLRAGGAEHRTLSGNKAGQWAPGEGST 60  
61 AFKCLLSARLCAALLSNISDCDFTFNWEPHYLIYEGGFTWYSPAYAIRSYALLH 120  
61 AFKCLLSARLCAALLSNISDCDFTFNWEPHYLIYEGGFTWYSPAYAIRSYALLH 120  
121 AWPAAFHARILQNKILVFFLRCCLAFVSCICELYFYKAVCKKGLHVSRLMFLVLS 180  
121 AWPAAFHARILQNKILVFFLRCCLAFVSCICELYFYKAVCKKGLHVSRLMFLVLS 180  
181 TGMFCSSAFLPSFCMYTTLIANTGMYDKTSIAVLGVAAGAILGWPFSAALGLPIAFD 240  
181 TGMFCSSAFLPSFCMYTTLIANTGMYDKTSIAVLGVAAGAILGWPFSAALGLPIAFD 240  
241 LLVGMHRKWSFFHSLMALILFLVPVVVVIDSYTGKVIAPLNIIVLVNFTPHGPDLYGT 300  
241 LLVGMHRKWSFFHSLMALILFLVPVVVVIDSYTGKVIAPLNIIVLVNFTPHGPDLYGT 300

301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVNGLGHPYWLTLAP 339  
301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVNGLGHPYWLTLAP 339  
RESULT 3  
US-09-922-225A-6  
Sequence 6, Application US/09922225A  
Publication No. US20030104385A1  
GENERAL INFORMATION:  
APPLICANT: Evans, Glen A.  
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
FILE REFERENCE: P-EA 4672  
CURRENT APPLICATION NUMBER: US/09/922,225A  
CURRENT FILING DATE: 2003-01-14  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 331  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-922-225A-6

Query Match 45.5%; Score 1489.5; DB 11; Length 331;  
Best Local Similarity 89.3%; Pred. No. 4.9e-136;  
Matches 284; Conservative 5; Mismatches 14; Indels 15; Gaps 3;

295 PDLYGTETFWFYLLNGFL-NPNVAFALALLVPLTSLMEYLQRFHVNGLGHPYWLTLAP 353  
28 PVLYSKFW--KVDGGTQNFN-----SRTEFCMLCVQVQNLGHPYWLTLAP 73  
354 MYIMFIIPFIQPKHERFLFPVYPLICLCGAVALSALOKYHFVQRYRLEHYTVTSNWL 413  
74 MYIMFIIPFIQPKHERFLFPVYPLICLCGAVALSALOKYHFVQRYRLEHYTVTSNWL 133  
414 ALGTVFLFGLLSFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWY 473  
134 ALGTVFLFGLLSFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWY 193  
474 RPPSSFLIPDNWQLOFTPSBERGOLPKPFAEGPLATRIPTDMDNDQNLLEEPSRIDISK 533  
194 RPPSSFLIPDNWQLOFTPSBERGOLPKPFAEGPLATRIPTDMDNDQNLLEEPSRIDISK 253  
534 HYLVDLTWRETTPREPKYSNKEEWSLAYRPFLLDASRSSKLLRAFYVFFLSDQYTVVYN 593  
254 HYLVDLTWRETTPREPKYSNKEEWSLAYRPFLLDASRSSKLLRAFYVFFLSDQYTVVYN 313  
594 YTIKPRKAKQIRKXSGG 611  
314 YTIKPRKAKQIRKXSGG 331

RESULT 4  
US-10-050-704-121  
Sequence 121, Application US/10050704  
Publication No. US20030050442A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
TITLE OF INVENTION: 62 Human Secreted Proteins  
FILE REFERENCE: P2039P1  
CURRENT APPLICATION NUMBER: US/10/050,704  
CURRENT FILING DATE: 2002-01-18  
PRIOR APPLICATION NUMBER: 09/684,524  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: PCT/US00/08979  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: 60/128,693  
PRIOR FILING DATE: 1999-04-09  
PRIOR APPLICATION NUMBER: 60/130,991  
PRIOR FILING DATE: 1999-04-26  
NUMBER OF SEQ ID NOS: 344  
SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 121
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-050-704-121

Query Match
Best Local Similarity 42.6%; Score 1395; DB 15; Length 258;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 MYWFIIFQPHKEERFLPVYPLICLCGAVALSALQKCYHFVQRYRLEHYTVTSNKL 413
Db 1 MYWFIIFQPHKEERFLPVYPLICLCGAVALSALQKCYHFVQRYRLEHYTVTSNKL 60

QY 414 ALGTVFLGGLSFSRSVALFRGYHGPLDLYPEFYRIATDTTHTVPEGRVNVVCVKWY 473
Db 61 ALGTVFLGGLSFSRSVALFRGYHGPLDLYPEFYRIATDTTHTVPEGRVNVVCVKWY 120

QY 474 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLPEPSRYIDISK 533
Db 121 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLPEPSRYIDISK 180

QY 534 HYLVDLDTMRETPREPCKYSNKEEWSILAYRPFDDASRSSKLLRAFYVFPFLSDQYTYVN 593
Db 181 HYLVDLDTMRETPREPCKYSNKEEWSILAYRPFDDASRSSKLLRAFYVFPFLSDQYTYVN 240

QY 594 YTIKPRKAKQIRKXSGG 611
Db 241 YTIKPRKAKQIRKXSGG 258

RESULT 5
US-09-922-225A-14
; Sequence 14, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 623
; TYPE: PRT
; ORGANISM: D. melangaster
US-09-922-225A-14

Query Match
Best Local Similarity 41.5%; Score 1358; DB 11; Length 623;
Matches 267; Conservative 115; Mismatches 213; Indels 34; Gaps 10;

QY 1 MASRGARQRLKSGSGSDGTAPADKRLRLG---SREAGAEHRTLSGNKAGQVWA-- 55
Db 1 MAPPAARARYIAKADNOILSKPPRPLGNKNKTEATPAGKKDKADKKNQPTSGG 60

QY 56 -----PEGSTAFKLLSARLCAALISNIDCDETFNWEPTHTLYLVEGQFQWY 105
Db 61 QKGLPNPMPVSQAKTFVFSARLCSAIWAYIADCDTFNWEPLHYIINGHGLQWY 120

QY 106 SPAYAIRSYAILLHAWPAAFHARILQTNKILVFLRCLLAFVSCICELYFKAVCKKF 165
Db 121 SFQGLRSYTYLLQGVPGYFYQKLFNPSILIFVWACMLGFGCAVMERYWYKSIQCEB 180

QY 166 GLHVSRMALFVLSTGWFCCSSAFPLPSSFCNWTTLIAMTGMWMDKTSIAVLGAACAIL 225
Db 181 GHIGRLWILFQVFGVMPVSSTALLPSSFMVFCALAAWQNVCFALFIAISALL 240

QY 226 GWPFSAALGLPIAFDILLVMKRWKSFHWSLMALILFLVPVVVDSYVYKGLVAPLNIY 285
Db 241 GWPFALIGLPIVLEMLLRQDWKTFVQWTLISGATVAIPMAIDTSYFGKLTAPLNIY 300

```

```

QY 286 LYNVFTPHGPDLYGTEPWYFYLINGELNENAFALALLVPLTSLMEYLLQRFHYQN-LG 344
Db 301 WYNVFTSHGNIFGTEPLSYFYIINGELNENIWLAL-LQLPIMLVIDLIVPAKSKSTLN 359

QY 345 HPYWLTLAPMYIWIIFFTQPHKEERFLPVYPLICLCGAVALSALQKCY----HFVQOR 400
Db 360 PPHYISLAPLYLWLVFFAQPHEERFLPPIYPLISLCGATVDVYVQRIFFRPMKSVFKI 419

QY 401 YRLEHYTVTSNWLALGTVFLGGLSFSRSVALFRGYHGPLDLYPEFYRI-AT---DPTH 456
Db 420 KAGVHYLDHSMFATILVMVTSTLLGSRVFLYRNTHAPMDLMLNQLKATPOYDQDV- 478

QY 457 TVPEGRVNVVCVKWYRPPSSFLP-DNQLOFIPSEFRGOLPKPFAEGPLATRIVPTD 515
Db 479 -----LYNVICIGKWHRYPGSFFPAKXFLRFLKSEFRGMLPAYYDEGNATKVQVY 532

QY 516 MNDONLEEPSRYDISKCHYLVDLDTMRETPREPCKYSNKEEWSILAYRPFDDASRSSKL 575
Db 533 FNDLNQENEHYFDYDCDFLVDDEGKYALEPNYSKRKSKOWSMKSLPFLIPEKSHKV 592

QY 576 LRAFYVPFLSDQYTYVYNYTILKPKAKQ 604
Db 593 LRAFYVPFLTDNHQYGDENLLK-RKTKR 620

```

```

RESULT 6
US-09-922-225A-4
; Sequence 4, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-225A-4

Query Match
Best Local Similarity 31.0%; Score 1013.5; DB 11; Length 255;
Matches 193; Conservative 5; Mismatches 14; Indels 15; Gaps 3;

QY 295 PDLYGTEPWYFYLINGEL-NFNVAFALALLVPLTSLMEYLLQRFHYQNIGHPYWLTAP 353
Db 28 PVLYSKFW--KVDGQGTQNFNL-----SRTEFCLWCVPQNLGHPYWLTAP 73

QY 354 MYWFIIFQPHKEERFLPVYPLICLCGAVALSALQKCYHFVQRYRLEHYTVTSNWL 413
Db 74 MYWFIIFQPHKEERFLPVYPLICLCGAVALSALQKCYHFVQRYRLEHYTVTSNWL 133

QY 414 ALGTVFLGGLSFSRSVALFRGYHGPLDLYPEFYRIATDTTHTVPEGRVNVVCVKWY 473
Db 134 ALGTVFLGGLSFSRSVALFRGYHGPLDLYPEFYRIATDTTHTVPEGRVNVVCVKWY 193

QY 474 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQ 520
Db 194 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDMNDQ 240

RESULT 7
US-09-922-225A-13
; Sequence 13, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides

```

TITLE OF INVENTION: Associated with Bipolar Disorder  
FILE REFERENCE: P-EA 4672  
CURRENT APPLICATION NUMBER: US/09/922,225A  
CURRENT FILING DATE: 2003-01-14  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 653  
TYPE: PRT  
ORGANISM: C. elegans  
3-09-922-225A-13

Query Match 30.1%; Score 986; DB 11; Length 653;  
Best Local Similarity 36.5%; Pred. No. 1.2e-86;  
Matches 217; Conservative 99; Mismatches 213; Indels 66; Gaps 13;  
Y 48 NKAGQVAPGEGTAKKLLSARLCAALLNSISDCDETFNWEPTHYLYYGEQFTWEYSP 107  
b 43 NNPDNDWPFSGVFMKLLSIRISGAIWGIINDCEVNTWYWEPLHLFLYGEQFTWEYSP 102  
Y 108 AVAIRSYAVLLHAWPAAPHARILQTNKILVYFLRCLLAFVSCICELYFYKAVCKKFG 167  
b 103 VVAIRSYFYIYHLYPASLPAFLPGDTKIVVFTLRLTIGLFCLLGEVYAFDAICKKINI 162  
Y 168 HVSMMKLAFLVLTGMPGSSSAFLPSSFCMTTLIAMTGMNMTKSTAVLGVAAGAILGW 227  
b 163 ATGRFFILFISFGGFLASTAFVPSFCAITFYILGAYLNENWTAGIFCVAFSTWVGW 222  
Y 228 PPSAALGLPIADLLVMKHWKSRFFHMSMALILFLVPVVVDSYVYKLVIAPLNIVLY 287  
b 223 PPSAVLGLPIVADMLLGLRIRFLTSLVGLGIGGVQVITDSHYFGKVLAPLNI 282  
Y 288 NVYTPHGPDLYTEPWYFYINGFNFNVAFAALLVPLTSLMEYLLQRFPHVQNGHPY 347  
b 283 NVYSGGSPSLYGEELPSFYIKLNNMNIIFAAPFGPLS--LAYFTKVMMSQDNVAL 340  
Y 348 WLTLAPMYI-----WFTIFPIQHKERFLFPVYPLICGGAVALSALQCKVHFVFOR 400  
b 341 YQRFAPITLLAVTTAAWLLIFSGQAKKERFLFPYIPFIAPFAALADATNR---LCLKK 397  
Y 401 YRLHYVTNSNWLALGTFLGLLSRSVALFRGYHGLDLYPEFYRIATDPT-----IH 456  
b 398 LGWD-----NLSILFLICFAILSASTYSIHNNYGSVHEIYRSLNAELNTRNFKNFH 451  
Y 457 TPEGEPVNVGVCKEYRPPSSFLPD-----NWQLQFIPSEFRGOLPKPEAGPL--- 507  
b 452 D-----PIRVCVKGWHRFPSSFFIPQTVSDGKKVEMEFIQSEFRGLLPKPELSDKLVE 506  
Y 508 ATRIVPTMNDONLEBPSRYIDISKCHYLVDLTWRETREPCKYSSNKEE----- 557  
b 507 VTRHIPTENNLNQEEISRYVLDSDYVVD-MPQSDREPDFRQKQNYKRTKKWK 565  
Y 558 -----WISLAYR--PFLDASRSSKLLRAFYVFFLSDQ---YTVVYNTILK 598  
b 566 RIENAILQFWLNSLFEIKFLNSN-----FHSFPFFVFDNFGLFQAYSHYSFLK 615

RESULT 8

3-09-922-225A-11  
Sequence 11, Application US/09922225A  
Publication No. US20030104385A1

GENERAL INFORMATION:  
APPLICANT: Evans, Glen A.  
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
FILE REFERENCE: P-EA 4672  
CURRENT APPLICATION NUMBER: US/09/922,225A  
CURRENT FILING DATE: 2003-01-14  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 577  
TYPE: PRT

ORGANISM: S. cerevisiae  
US-09-922-225A-11

Query Match 29.7%; Score 973; DB 11; Length 577;  
Best Local Similarity 37.7%; Pred. No. 1.8e-85;  
Matches 219; Conservative 94; Mismatches 204; Indels 64; Gaps 18;  
QY 53 VWAPGEGTAFKCLLSARLCAALLNSISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIR 112  
Db 15 VW-----TESILAVLRLTSASPRVIDDCDEVYNYWEPLHYLLYGYGLQTWESPEYAIR 68  
QY 113 SYAVLLHAWPAAPHARILQTNKILVYFLRCLLAFVSCICELYFYKAVCKKFGFLHVSRL 172  
Db 69 SWFVIALHAYP-QFLARGGLSLRHLVYFIRGVLACFSACFCTNLIILAVARNFNRAVALH 127  
QY 173 MLAFVLSTGMPGSSSAFLPSSFCMTTLIAMTGMW---MDKTSIAVLG-VAAGAILGWP 228  
Db 128 LTSVLVFNQGWASATSFLESSFAMNVLIALSAQLSPSTKRTKVVSFITIGAVIGWP 187  
QY 229 PPSAALGLP-IAFDLLVMKHWKSGFF-HWSLMALILFLVP--VVVIDSYVYKLVIAPLNI 284  
Db 188 FSAALSIPFILLEVLVDLKGFRHLFCRWFKAFVALLITGICITVDSLFYHRIQFVAMNI 247  
QY 285 VLVNVFTP--HGPDLYTEPWYFYINGFNFNVAFAALLVPLTSLMEYLLQRFHVQN 342  
Db 248 VKNVLADGEGDPIYGTETPWYFYFANLSLQHNIVLWFAWACGPLVLAAP----- 298  
QY 343 LGHPYWTLL-----APMYINFIIPFIQHKERFLFPVYPLICGGAVALSALQK 392  
Db 299 ---TNWINDSFLDSSVISGPFYIWLFIPIQHKERFMYPIYPLVCLAAAIGLDMSLK 355  
QY 393 CYHVFORVRLHYVTVS---NWLALGTFLGLLSRSVALFRGYHGLDLYPEFYR 448  
Db 356 ---LMIQLLSINETVRSKFPVRFVLCVTAIIGCLSLIARLAI-QYNAPMIYPAISF 411  
QY 449 IATDPTIHTVPEGRPVNVGVCKEYRPPSSFLPDNWQLQFIPSEFRGOLPKPEAGPL- 507  
Db 412 LETDNVTT-----NVCVKGWYRYPSTFFLPDNRSLKFKVSEDFGILPGEFVESNST 464  
QY 508 -----ATRIVPTMNDONLEBPSRYIDISKCHYLVDLTWRETREPCKYSSNKEEWS 560  
Db 465 WNNREGYVQIPEHNNFNEEPTRYLSLESCDFLIDLEFDHRSKATNNEPIYSKS-DGWIP 523  
QY 561 LAYRPFIDASRSSKLLRAFYVFFLSDQYTVVYNTILKPKK 601  
Db 524 VWYVFFIDTKTPEWGRAFAVPFTEPKWGRY-EILVKKPVK 563

RESULT 9

US-09-922-225A-15  
Sequence 15, Application US/09922225A  
Publication No. US20030104385A1

GENERAL INFORMATION:  
APPLICANT: Evans, Glen A.  
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides  
FILE REFERENCE: P-EA 4672  
CURRENT APPLICATION NUMBER: US/09/922,225A  
CURRENT FILING DATE: 2003-01-14  
NUMBER OF SEQ ID NOS: 117  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 577  
TYPE: PRT  
ORGANISM: S. pombe  
US-09-922-225A-15

Query Match 29.7%; Score 973; DB 11; Length 577;  
Best Local Similarity 37.7%; Pred. No. 1.8e-85;  
Matches 219; Conservative 94; Mismatches 204; Indels 64; Gaps 18;  
QY 53 VWAPGEGTAFKCLLSARLCAALLNSISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIR 112

```

15  VW-----TF8ILAVRLTSASFVIDDBEVVNNWEPHLHYLYGLOTWTSYBEYAIR 68
113 3YAYILLHANPAAFEARILQTNKILVFLYRCLLAFVSCICELYFYKAVCKKFGFLHVRM 172
69  SWFYIALHAVP-GFLARGLGLSRHLVFFYIRGVVLACFSAPFCETNLILAVARNFNRAVALH 127
173 MLAFVLVSTGFCSSSAFLPSSFCWYITLTAWTGWY---MDKTSIAVLG-VAAGAILGWP 228
128 LTVSLVFNMGWSASTSLPSPFANNWVTLALSAQAQSPSPSTKTKVSVBFTTIGAVIWP 187
229 FSAALGLP-IAFDLLVMKHWKSPF-HWSLWALILFVP--VVVIDSYYGKLVIAPLNI 284
188 FSAALSIPFILLLELDVLRGRPHLFCWFKRAIFVALLITGICTVDSLFSVHRIQFVANNI 247
285 VLYNVYFTP--HGPDLYGTEPVYFYLINGFNFNVAFALLVLLPLTSLMEYLLQRPHVN 342
248 VKYNLWAKDGRGPDLYGTPEWYIYFANLSIQHNILWFMACGLVLLAAF----- 298
343 LGHPYWLTL-----APMYIWFITFIQPHKEERFLFPYYPILIC:CGAVALSALQK 392
299 ---TWKINLSPDLSSVISPFYIWLFTFIQPHKEERFMYPYIYVPLCAAIGIDMSLK 355
393 CVHFYFQYRVLHYHTVS---NWLALGTVFLGLSFSRSVALFGYHGPDLDLXPEFYR 448
356 ---LMIQILSSINETVRSKPPVRFVLCVTAITGLSIRALAI-QYNAPMIIIPATSF 411
449 IATDPTIHTPEGRPVNVYCVGKEWYRPPSGFLPDNNQLQFIPEFRRGOLPKPPFAEGPL- 507
412 LETDNNVTT-----NVYCVGKEWYRPPSTFFLPDNSRLKFKVASEPDGILPGEFVESNST 464
508 ----ATRIPTDMNDQNLBPSRYIDISKHYLVDL--DTMETPREPKYSSNKEEMTS 560
465 WNNRGYVQIPEHMMFNFNEBPTRYTLESCDFIDLEFDHSHKATVNEFYSKS-DGWIIP 523
561 LAYRPFLDASRKSLLRAYFYVPFLSDQYTVVNVYITILKPRK 601
524 VNVYFPIDTKOTPFNGRAFAFYFTEPKWGRY-ELIVKKPVK 563

```

```

RESULT 10
JS-09-922-225A-12
; Sequence 12, Application US/09922225A
; Publication NO. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; TITLE OF INVENTION: Associated with Bipolar Disorder
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 570
; TYPE: PRT
; ORGANISM: A. thaliana
US-09-922-225A-12

```

```

Qy 184 FCSSAFIPSSFCWYTTLIANTGMYDKTSIAVLGVAAAGAILGWPPSAALGLPIAPDILIV 244
Db 172 FFASTSFLPSSFSMYA:SLSGLLLPFKYAMAVAVSVGVILGWPFESILAFPLFWIYSLV 231
Qy 244 MKHRWKSFFHWSLMALILFLVPVVVIDSYYGKLVIAPLNIVLVNFTPHGDPDLYCTEPW 303
Db 232 --KRFKQAFIAGAVTTIFLLGVSLLVDYVYKRWTSVSNLLIYVVGGSHELHYTEGA 289
Qy 304 YPYLINGFLNFVAFALALLVPLTSLMEVLLQRFHVQNLGHYPWLTLPAPMYITFIIFI 363
Db 290 LEYIRNGNNFNFCFILALVALYPIVRKYDR-----ALLVVISPMYIWLAFMSL 341
Qy 364 QPKKERELFPYPLICLCGAVALSALQCKYCHFVORYRLEHYTTSNNMALGTWFLFGL 423
Db 342 QPKKERELFYPLIYCVASAVTENIPELFR-----EKYSSESRESLVTITKVMRPV 393
Qy 424 L-----SPFSRVALFRGVHGPLDLYPEFVRIATDPTIHTVEGRPVNVCVGKEWYRP 476
Db 394 ILCILCHASHERTFALINGSAPLEVYKLE-----HHDDAGPGSVLCVGSEWHRYP 445
Qy 477 SSFLPLDN-WOLQTFPSFRQLPKPFAEGPLATRIVPTMDNQNLLEBSPRYI-DISKCH 534
Db 446 SSSFFVPHYSISVRNIDGFRLLPFPFNNTLGTSASPPYFNKNQOASBEQVLKNIETCT 505
Qy 535 YLVLDVDMREPREBPKYSNKEEMISLAYRPFIDASRSSKLLRAFYVVPFLSDQYTVVNY 594
Db 506 FLIEQLSR--PYQR-GSDLSLWETAVILPYLDRELSPAKYRSFFIPHMWQEKNVFGKY 562
Qy 595 TIIK 598
Db 563 VALR 566

RESULT 11
US-10-291-265-240
; Sequence 240, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 240
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-240

```

```

Query Match      6.4%; Score 209.5; DB 12; Length 554;
Best Local Similarity 19.9%; Pred. No. 3e-11;
Matches 109; Conservative 85; Mismatches 196; Indels 159; Gaps 26;

Qy      87 YW---EPHYATLYEGEGQTWEYSAPAXIRSAIVAYLLHLHAPAAFHARILOTKNILVFVFLR 143
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      87 YQMSLEVSHWVFNFGYLTWEITE--RLRSYTYPLFIASIIYKILHLLGKDSVQLLIWIPR 144
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      144 CILAFVSIGTICELYFKAVCKCKFLGHVSRMMFLAFVLSTGMFCSSAFIPSSFCMYT---- 199

bb      145 LACALLISAVDRLYLSMLKOLENOEVARWY-----FFCOLCSWFETWYCCTRLTNT 195
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

200 -----TLIANTGWYND-----KTSIAVLGVAAGAILGWPFSAALGLPIAFDPLLVMKRWKS 250  
 196 METVLTITIALFYPLEGSKSMNSVKYSSSLVALAIIIRPTAVILWTLPLFRHFCQEP-- 252  
 251 PFHMSMALILFLVPV-----VWDSYTYGKLVIAPLNVLNVTTPHGPDLGYTEP 302  
 253 -----KDLILHFLPVGVVTVLSLMIDRIFFGQWTLVQFNFLKFNVLQWNG-TFYGSH 307  
 303 WYFYLYNGS---LNFNVAPALALLVPLTSLMEYLLQRFHVNQNGHYPWTLAPMVIW-F 358  
 308 WHWYFSQGFVILGTHLPEFI-----HGYLAPKRYRILLVTLVTL 349  
 359 IIFFIQPHKEERFLPVVFLICLGCAGVALSALOKCYHFVQRYLEHYTVTSNWL--ALG 416  
 350 LVISMLSHKEFFIIPVLP-FCM-----VFCGYSLTH--LKTWKRPALS 390  
 417 TVPLFGLLSFSRSVALFRG---YHGPDLYFEFVRIATDPTIHTVPGRPVNV-----CV 468  
 391 FLPLSNLF-----LALYTLGVHQRGTLDVMSHIQKVC-----YNNPNKSSASIFIMPC 440  
 469 GKEMWFFPSFLLPDNWLQPTPSEFRGOLPKPFAEGPLATRIVPTDMNDQNLSEPSYI 528  
 441 STPIY-----SHVHCLPMRFLO-----CPDILT----- 464  
 529 DISKCHYLDLMTRETPREPKYSNKEEWISLAYRPFELASRSKLLRAFVYFFLSDQY 588  
 465 --CKSHYLDADV-----FYLNLNLW---HREFHDDASLPTHLITFSI--LBEI 508  
 589 TVYV---NY 594  
 509 SAFPSSNY 517

RESULT 12  
 S-10-080-170-547  
 Sequence 547, Application US/10080170  
 Publication No. US20030129601A1  
 GENERAL INFORMATION:  
 APPLICANT: COLE, S.T.  
 TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
 IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
 TREATMENT OF MYCOBACTERIOSES  
 FILE REFERENCE: 03495.0218  
 CURRENT APPLICATION NUMBER: US/10/080,170  
 CURRENT FILING DATE: 2002-06-10  
 PRIOR APPLICATION NUMBER: 60/270,123  
 PRIOR FILING DATE: 2001-02-22  
 NUMBER OF SEQ ID NOS: 652  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 547  
 LENGTH: 433  
 TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis  
 S-10-080-170-547

Query Match 4.0%; Score 130; DB 16; Length 433;  
 Best Local Similarity 21.5%; Pred. No. 0.0011;  
 Matches 103; Conservative 62; Mismatches 169; Indels 144; Gaps 24;  
 18 GDTAPADKURELLGSRAGGAHRTLSGNKAGQWABEGSTAFKLLSARLCHALLS- 76  
 3 GALVTAADSIRTGLGASLLAGRPRT-----GAPSTAT---ILRSALWPAANLSV 49  
 77 -----NTSD-----CDTFNWEPTHYLIYGEFQWTEYSP 107  
 50 LHSIVLTNGNITDDPKPYRAVLNFRGWDIYNEHFDYDP-HYL-----YPP 98  
 108 AYAIRSVAYLLHLAWPAAPFARILQNKILVFLFCLLAFVSCICELYFKAVCKEGL 167  
 99 GGTLLMAFFGLFPAPSRYLFIISNTAAILVAYL-----LLRMFNF 140  
 168 HVSRMMLAFVLSTGMFCSSS-----AFLPSSPCMYTTLIAMTGWMDKTSIAVLGVAAG 222

141 TLTSAAPALILA--MPATETVNTVFTNINGCILLLEVLRLWLD-----GR 188  
 223 AILGWPFSAALGL-----PIAFDLL---VWKRHWKSFFHWSMALILFLVPVVDSDYY 274  
 189 ASRQWCGGLAIGTLVLKPLGPELLPLINRQR-----ALVAAV--VVPVV 235  
 275 GKLVIAPLNVLNVTTPHGPDLGYTEPWFYLYNGFLNFNVAFA-----LALLVLP 326  
 236 --NVAALPLVSDPMSFFTRTLPYLGTED---YFNSILGNVYFGLPTWLLILFLRLFTA 291  
 327 LTSMEYLLQRFHVNQNGH-PYWTIAP-----MYIWIIFFIQPHKEERFLPVVFLICL 381  
 292 ITFGALMLLYRY--RTGDPLEWFTTSSGVLILWSLWMSLAQGY-YSMMLFPFLMTVVL 348  
 382 -----CGAVALSALOKCYHFVQRY--RLBHYTVTSNWLALGTVFLFGLLSF 426  
 349 PNVIRNWPALVGVYGFMTLDRWLLFNWNGRALBYLKITYGM-SLLIVITVTVLYF 405

RESULT 13  
 US-10-032-585-7898  
 Sequence 7898, Application US/10032585  
 Publication No. US20030180953A1  
 GENERAL INFORMATION:  
 APPLICANT: Terry, Roemer D.  
 APPLICANT: Bo, Jiang  
 APPLICANT: Charles, Boone  
 APPLICANT: Howard, Bussay  
 TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
 FILE REFERENCE: 10182-005-999  
 CURRENT APPLICATION NUMBER: US/10/032,585  
 CURRENT FILING DATE: 2001-12-20  
 NUMBER OF SEQ ID NOS: 8000  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 7898  
 LENGTH: 584  
 TYPE: PRT  
 ORGANISM: Candida albicans  
 US-10-032-585-7898

Query Match 3.8%; Score 126; DB 12; Length 584;  
 Best Local Similarity 21.2%; Pred. No. 0.0042;  
 Matches 97; Conservative 69; Mismatches 149; Indels 142; Gaps 25;  
 198 YTTLIAMTG---WYMDKT-----STAVLGVAAG---AILGWPFSAALGLP 236  
 139 YTTLLSQPHLYTSSRTLPNFIALPLVNFSLKLIQGLSLGTWLAFTGIVFRLEVGL- 197  
 237 IAFDILLVMKRWKSFFHWSMALILFLVPVV-----IDSYYGKLVIAPLNVLN 288  
 198 --FGLIIAIVSSLSGFGQSNIFGNIYILAMGTLLGGT\*SFICIDSYFWMGRPLIPEIDSIFN 255  
 289 VFTPHGPDLYGTEPW--YF--YLINGFLNFNVAALALLVPLTSLMEYLLQ--RF-HV 340  
 256 IVQGSTK-WGTEPMDTYFKKYL-----FQLEPPVILMLAIPGLINDPANDGTKEGDK 308  
 341 QNLGHPYMTLAPYIWIIFFI---IQPHKEERFLPVVFLICL-----CGAVAL 387  
 309 KSVPHAPARYSLNLFISILFIAAMSFPQHKWRFIVVTIPIT\*LOAANGVTNICQKWL 368  
 388 SALQKCYHFVQRYLEHYTVTSNWLALGTVFLGLLSFSRSVALFRYHGPDLGYPEFY 447  
 369 SVLNKVLFIIG-----ANVTISLLS-----LHWYISSFN-YF--- 402  
 448 RIATDPTIHTVPEGRPVNVCVKWEYRFPSSFLPDNWLQF-----IPSEFRGOLPKP 501  
 403 -----GGDALQFTNNVIL-ENYKNETVSVHMDVPACMTG----- 435  
 502 PABGPLATRIVPTDMNDQNLSEPSYIDISKCHYLDLMTRETPR-EPKYSSNKEEWIS 560  
 436 -----ITRFGELDKGFASYDKSEQDFDITNYDIIII---THNEVFNWELLHSSRVFGIS 486  
 561 LAYRPFELASRSKLLRAFVYVFLSDQYTVVYNTIL 597



106 487 L-RMFIQIFMAOR-----KDRSHIP-NITKL 510

RESULT 14

US-10-080-170-164

Sequence 164, Application US/10080170

Publication No. US20030129601A1

GENERAL INFORMATION:

APPLICANT: COLE, S.T.

TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR

TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR

TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES

FILE REFERENCE: 03495.0218

CURRENT FILING DATE: 2002-06-10

PRIOR FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 652

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 164

LENGTH: 440

TYPE: PRT

ORGANISM: Mycobacterium leprae

US-10-080-170-164

Query Match 3.6%; Score 117; DB 16; Length 440;

Best Local Similarity 20.6%; Pred. No. 0.021;

Matches 80; Conservative 59; Mismatches 146; Indels 104; Gaps 20;

106 82 DETENYWEPTHYLIYEGFQWTSYPAIRSYAYVLLHAWPAAPHARILQTNKILVEYF 141

106 84 NEHPDYVDP-HYL-----YPPGTLMAFPGLPAPSRVLFILINTGALLIAWY 132

106 142 LRCILAFVSCICELYFYKAVCKFGHVSRRMLAFVLSTGMFC-----SSSAFLPSSFC 196

106 133 L-----ILRLPKVTLSSVAAPTLLA--MFCETVTSTLVTFTNGC 172

106 197 MYTTLIANTGWMKTSIAVLGVAAGAILGWPPSAALGL-----PLAFDL--VMKRW 248

106 173 IMLEVLFLMWLINGSFETKVSQO-----WAGGAIGTLVLKPLGLPLCLPLLNQW 226

106 249 KSFFHWSMALI-LFLVPVVVIDSYVYKGLVIAPLNVLNVFTPHGPDLYGTEFW---- 303

106 227 Q-----ALVPAIALPVVI-----NLAAPLVSHPMDFTRVPYILGTRDYFNSS 271

106 304 -----YFYLINGELNVAFAFALLVPLTSLMEYLLQRFVQNLGHPYWLTP----- 353

106 272 IEGNGVYFGLFTWLIVF---LRLFTVLAICSL--WLLARYY-RTDRPLFWFTCTGVLL 325

106 354 MYIWFIFFIQPHKEERFLFPVYPLICL-----CGAVALSALQCYHFVFORQY- 401

106 326 LWSLVLPLAQGY-YSNMLFFFLMTVTVLNSLIRNWPAMLIYGLTLDRLWLLFNWRYG 384

106 402 -RLEHYTVTSNW-----LALGTVFLFGLL 424

106 385 RALEYKITYGWSLLIIVSVTLCFRYL 412

Search completed: January 13, 2004, 07:09:54

Job time : 56 secs

106 105 YSPAVAIRSYAYVLLHAWPAAPHARILQTN--KILVYFLRCLLAFVSCICE 154

106 116 FNPCEIDMGYNILRVLIWFISITAITGNIIVLVIITTSQYKLTVPFLMCLAFAD-LC- 173

106 155 LYFYKAVCKFGHVSRRMLAFVL--LSTGMFCSSAF--LPSSFQCMYT-TLIAMTGWY- 208

106 174 IGIYLLLIASVDIHTKSYQHNVAIDWQIGAGCDAGFTVFASLSVYTLTALIERWHT 233

106 209 -----MDKTSIAVLGVAAGAILGWPPSAALGLPTAFDLVVMKRWKSFHWSL----- 256

106 234 ITHAMQLD-CKVQLRHAASVMVMGWIFAFAAALPFIFGI-----SSYMKVSICLPMDI 285

106 257 -----WALILF--LVPVVVIDSYVYKGLVIAPLNVLNV-----VFTPH 293

106 286 DSPLSOLYMSLLVLNVLAFAVVICGCIHIYLTVRNPNIVSSSDTRIAXKAMLIFF-- 343

106 294 GPDLYGTEFWYFYLINGELNVAFAFALLVPL--TSLMEYLLQRFH-VQNLGHPYWL 350

106 344 --DFLCWAPISFFAIS-----ASLKVPLITVSKAKILLVLPHPINSCANPFLYA 390

106 351 LAPMYIWFIFFIQPHKEERFLFPVYPLICLGAVALSALQCYHFVFORVLESHYTVTS 410

106 391 I-----FTKNFRDRDFFI-----LLSKCG-----CYEMOQAIYRTTSTSTVH 426

411 N 411

427 N 427

Search completed: January 13, 2004, 07:09:54

Job time : 56 secs